

# Evaluating Public Research and Innovation Policies: A Short History of Co-evolution



Susanne Bühner, Jakob Edler, Stefan Kuhlmann, and Sarah Seus

**Abstract** In this article, we present the mutually development of R&I policies and R&I policy evaluation with a specific focus on Germany and the EU. We show that there has been considerable development in the field of R&I policy evaluation over the past decades, in terms of underlying theories, methodological approaches and the emergence of an R&I evaluation community. The field of R&I policy evaluation has naturally interacted and grown in parallel with the changes in R&I policy and practice as well as scientific advances. In line with the actual claims of R&I policy to support transformation, R&I policy evaluation has engaged in a debate around more responsible and inclusive evaluation procedures. The discussion about “responsible metrics” and societal stakeholder engagement in evaluation studies are examples of it. Our overview of evaluation practice has shown that the R&I policy evaluation community need continuous, in-built critical reflection on the (incessantly changing) role of evaluation in R&I policy.

---

S. Bühner (✉) · S. Seus

Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, Germany  
e-mail: [Susanne.Buehrer-Topcu@isi.fraunhofer.de](mailto:Susanne.Buehrer-Topcu@isi.fraunhofer.de); [sarah.seus@isi.fraunhofer.de](mailto:sarah.seus@isi.fraunhofer.de)

J. Edler

Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, Germany

Manchester Institute of Innovation Research (MIOIR), University of Manchester, Manchester,  
United Kingdom  
e-mail: [jakob.edler@isi.fraunhofer.de](mailto:jakob.edler@isi.fraunhofer.de)

S. Kuhlmann

University of Twente, Enschede, The Netherlands  
e-mail: [s.kuhlmann@utwente.nl](mailto:s.kuhlmann@utwente.nl)

© The Author(s) 2024

J. Edler, R. Walz (eds.), *Systems and Innovation Research in Transition, Sustainability and Innovation*, [https://doi.org/10.1007/978-3-031-66100-6\\_5](https://doi.org/10.1007/978-3-031-66100-6_5)

## 1 Introduction

The objective of this chapter is to discuss the interplay between the major lines of Research and Innovation (R&I) policy<sup>1</sup> and the role of explicit evaluation exercises that aim to capture the function and influence of such policies. Obviously, evaluation concepts and instruments correspond with the respective “fashions” and waves in R&I policy, such as cluster policies or competence centres at the end of the 1990s, the new mission orientation of recent years but also, with regard to institutional funding, the introduction of New Public Management approaches in public research organisations. Also partly considered in our chapter are the goals, instruments and approaches to evaluating research institutions. In this chapter, we focus on the German and the European level with regard to evaluation practice, however, with regard to evaluation theory we also include the Anglo-American literature (Guba and Lincoln 1989; Mayne 2001; Patton 1997; Scriven 1991; Weiss 1997).

The chapter is structured along a chronology of decades. For each decade, we describe core features and whether or how they differ from previous decades. The individual sections are subdivided according to the following aspects: (a) political ambitions and developments; (b) main aims and topics of evaluation; (c) applied methods and concepts and (d) emergence and development of an R&I evaluation community.

### 1.1 *What Are We Talking About? Some Definitions*

The German Evaluation Society (DeGEval) understands evaluations as “the systematic examination of the quality or utility of an object of evaluation” (DeGEval—Gesellschaft für Evaluation e.V. 2016; cited from Kohlweg 2019, 5). According to Kohlweg (2019), an evaluation is characterised by “a comprehensible systematic procedure based on empirically obtained data, which contrasts it with everyday evaluation processes, by a transparent, criterion-driven assessment that is carried out against the background of a specific context of use (investigation of utility) or across the board (investigation of quality), which contrasts it with pure research studies, and by its applicability to different objects. The most important objects of evaluation include projects, policy measures and other interventions (programmes), organisations, products, and evaluations themselves (meta-evaluation)” (ibid.).

Evaluations can be distinguished according to the timing (ex ante, interim, ex post), the purpose (summative, formative), the object of evaluation (see above,

---

<sup>1</sup>If not otherwise stated, we use “research and innovation policy” as a generic term to cover the various policy activities aiming to foster science, technology, research and innovation activities (Boekholt 2010). The terminology and foci of those policies have changed over time, and we will refer to those changes as needed. We focus on policies by ministries of research, science, innovation or economic affairs, i.e. we do not cover policies in other ministries that may support or influence the generation and diffusion of innovation.

projects, programmes, organisations, but also whole innovation systems) and the content-related (impact) dimensions (scientific, economic, societal, political, systemic impacts). Evaluation and impact analysis can hardly be separated, since questions about intended and unintended effects are typically part of every (programme) evaluation. At the same time, other communities, concepts and methodological approaches have emerged in the field of impact analysis, and these will be taken into account in the following wherever it is useful.

## **2 Beginnings: From the 1960s and the Evaluation of Research to the 1980s with the First Comprehensive Approaches to Evaluating R&I Policies**

### ***2.1 Political Ambitions and Developments***

Starting in the 1960s, in Western Europe, North America and beyond, societal, economic and political actors increasingly perceived needs to reform and strengthen post-World War II structures of economies, to advance social and welfare policies, to expand and modernise education systems and to make political-administrative systems more effective. Across the “varieties of capitalism”—from liberal to coordinated market economies—such reforms assumed quite different shapes (Hall and Soskice 2001). Altogether though, “modernisation” became a key term, in industry, politics and academia, often closely related to the term “democratisation” (e.g. Lipset 1959). Governmental public policies were expected to function as key resources in modernisation processes, understood as “problem solving”. To effectively solve problems, governments and their agencies would have to engage with “political planning” (Friedmann 1987; a term only partly overlapping with the French political concept of “planification”, see, e.g., Crozier 1965, or with the socialist-soviet concept of economic planning, see, e.g., Mandel 1986).

In the 1960s to 1980s, most authors understood “political planning” as rational end-means and decision-making action by political actors or political systems (see Deutsch 1973; Etzioni 1968). The conception is essentially based on an economic understanding of rationality, according to which “policy problems are seen as instrumental in nature, and policy makers are seen as rational to the extent that they do the best they can to satisfy the combined welfare functions of those affected by their policies” (Schön and Rein 1994, 11). This understanding has also been developed and partially practiced in Germany since the late 1960s as a targeted state modernisation strategy (e.g. Naschold 1969; Krauch 1970). Political planning presupposed a special role for the state: the central political authorities are ascribed the ability to conceptually shape the social environment (Martinsen 1992, 53). On the basis of a thorough reform of the political-administrative system (in terms of competencies, decision-making processes, structural and operational organisation), equipped with intelligent information about current and foreseeable future social,

economic and technological developments, “political planning” should enable effective state control and problem solving. The necessary generation and processing of information should take on a scientific character: “... effective policy-making will have to be future-oriented and will increasingly depend upon medium- and long-range forecasting and policy-planning capabilities” (Mayntz and Scharpf 1975, 5).

This is the political and intellectual context where governments made increasing efforts to support education, science, research and technological innovation no longer mainly through institutional funding of related organisations (higher education; research) but additionally also through policy programmes with defined aims, timing and formats.

- *Higher education and research:* Since the early 1960s higher education (and related research capacity) has seen a worldwide expansion: “... growth patterns are similar in all types of countries, [they] are especially high in countries more linked to world society, and sharply accelerate in virtually all countries after 1960” (Schofer and Meyer 2005, 898). In the same vein, in Germany the diagnosis of an educational crisis (“Bildungskatastrophe”, Picht 1964) marked the beginning of an educational expansion and an era of reform in the education (and related research) system, continuing to this day.
- *Science and technological innovation in industry:* After World War II, science policy was based on the belief that publicly sponsored scientific research would result in beneficial social and economic outcomes (see the influential report *Science the Endless Frontier*, Bush 1945). In the late 1970s and throughout the 1980s, in Western European countries, certainly in Western Germany, the competitiveness of national industries in international markets was perceived as weak, not at least in view of the rising economic power of Japan, so governmental funding policies were launched to “push” technological innovation in research organisations and industry. From the mid-1960s, empirical studies criticised such supply-push policy concepts for underestimating the role of market forces in the innovation process; as a consequence, demand-pull models of innovation were supposed to guide policy designs. As from the late 1970s, an increasing number of innovation studies scholars have argued that innovation should be understood as a two-sided process involving complex interactions between supply and demand (e.g. Freeman and Soete 1997; Boekholt 2010; Molas-Gallart and Davies 2006). In the Federal Republic of Germany, the innovative capacities of small and medium-sized enterprises (SMEs) were considered a crucial backbone with respect to competitiveness and employment. Consequently, governments supported the research, development and innovation capacities in SMEs.

The more publicly funded support measures were taken by governments and their agencies, the more the request grew—by the public, parliaments, media—to legitimise the investments by properly examining their efficiency and effectiveness.

## 2.2 *Main Aims and Topics of Evaluation*

Since the 1960s, in Germany as in other industrialised countries, heterogeneous lines of evaluation concepts and methods in the field of R&I policy have been prevalent. They can be represented as a shell model (Kuhlmann 1998b, 86pp, 2009, 285p): A first shell relates to individual research achievements: the “core” is formed by peer-review procedures and later additional procedures for measuring the research performance of individual researchers and groups (bibliometrics, etc.) as internal scientific instruments for deciding on the allocation of funding in research institutions (Daniel 1989). The second shell consists of the evaluation of political support programmes that, in contrast to institutional support, pursue politically decided and specifically designed intentions. A third shell relates to the performance of research institutions or larger scientific areas, as they have been carried out since the 1990s, e.g. by the Science Council (Wissenschaftsrat) (see also below).

In particular the second shell, the evaluation of research and innovation programmes gained traction in the 1970s and 1980s. “A case in point is the long and fruitless discussion among economists about the relative impact of direct and indirect government measures for promoting R&D on the national economy and technical and industrial development. This debate is mainly carried out under aspects of principles of political control and economic order, with little empirical backing and verification” (Meyer-Krahmer 1981, 358). Since the mid-1970s, however, research has been carried out by Fraunhofer ISI (Bräunling and Harmsen 1975) that has tried to overcome this “abstract either/or type argument and showed that the reality of government funding programmes must be viewed in a much more differentiated way” (Meyer-Krahmer 1981, 358).

## 2.3 *Applied Methods and Concepts*

The reform’s hopes conveyed by political planning efforts, on the one hand, and increasing disappointment about unintended effects of planning and programming, on the other hand, (Pressman and Wildaysky 1984) triggered a growing interest in understanding the actual *implementation* processes of programmes in bureaucracies vis-à-vis societal stakeholders. In Germany, a research project group “Implementation of political programmes” around Renate Mayntz (Mayntz 1979) studied basic aspects of policy implementation, both with internal administrative enforcement problems and with the difficulties of carrying out political programmes with the help of intermediary organisations. It became clear that effective policy implementation needs to understand, also from a methodological point of view, issues of norm formation, the effect of different types of programmes (regulatory policies, incentive programmes), the calculus of action at the local level in the implementation process, the coordination problems in the implementation process, etc. (Mayntz 1980).

Following a similar reasoning, colleagues at Fraunhofer ISI developed novel concepts and methods for the *evaluation of impacts* of governmental research and innovation programmes, in particular those providing R&D personnel cost grants to SMEs. With the help of surveys, interviews and statistical analyses the evaluation studies addressed the (1) acceptance of the funding by SMEs and the relevance to their internal decision-making; (2) quantitative and qualitative effects on R&D personnel; (3) impact on innovative activities of firms; (4) impact on the economy as a whole, such as employment, stimulation of innovation and diffusion processes and (5) efficient implementation and application of the measure. At the same time, the studies considered consequences for the governmental funding strategy, in the short and long term (Meyer-Krahmer 1981, 359).

Since the mid-1970s, in parallel with implementation and impact studies, the interest in *research and technology indicators* has grown sharply worldwide. Numerous studies have suggested indicators and used them to describe the level of performance in an inter-individual, inter-institutional and international comparison. While in early years the focus was mainly on input indicators for research and technology, in the 1980s the interest in result and impact indicators increased. Daniel (1989) provides an overview of the status and prospects of this field of science and innovation research, including an extensive bibliography of bibliometric, scientometric (e.g. Moed et al. 1985; Weingart et al. 1988) and technometric literature (e.g. Grupp and Hohmeyer 1986), the latter exploring the technological competitiveness of national economies, as an input in research and innovation policymaking and programme design.

#### ***2.4 Examples of Typical Research and Innovation Policy Evaluation Studies***

Throughout the three decades, first slowly then accelerating, a growing number of research and innovation (policy) evaluation studies were conducted. Daniel (1989) gives an overview of science-related evaluation efforts, Becher and Kuhlmann (1995) and Holland and Kuhlmann (1995) present numerous cases of technology and innovation policy evaluations in Germany. In the following, two typical examples are presented.

Fraunhofer ISI became a forerunner in Germany and Europe in the evaluative study of the actual impacts of SME research and innovation programmes. Drawing on conceptual work by Bräunling and Harmsen (1975) and Meyer-Krahmer (1981), in the early 1980s, Meyer-Krahmer et al. (1983) on behalf of the federal government conducted a seminal series of evaluation studies of the largest West-German government programmes offering grants to cover the costs of R&D personnel in SMEs. Subjects of investigation were “the quantitative and qualitative changes in R&D activities caused by the programme in the firms ... and the contribution the programme can make to strengthen innovative and competitive capabilities”

(Meyer-Krahmer et al. 1983, 153). Also the problem assumptions underlying the programme were reviewed critically, in particular to better understand “whether the barriers to innovation and the weak spots assumed did in fact exist in the firms examined” (Meyer-Krahmer et al. 1983, 153). This was important for the political legitimization of public funding; at the same time genuinely empirical insights into the innovation processes in SMEs were generated—a crucial contribution to the emerging academic field of “innovation studies” (see, e.g., Fagerberg et al. 2013). The administrative management, “free rider” effects, and factors restricting impacts were investigated along with such questions as to whether firms located in peripheral regions were reached by the programme to the same extent as firms in conurbation areas. Furthermore, the reasons for non-participating firms were analysed.

In 1983, the German Federal Ministry of Research and Technology (BMFT) started a pilot project to promote technology-oriented start-ups (TOU). In order to support this, in particular to expand the information basis for the technology-political decision-making in this new funding field, a scientific monitoring project was launched in 1984 and conducted for several years by Fraunhofer ISI (Kulicke 1993; see also Holland and Kuhlmann 1995, 175–187). Research work was carried out with three functions: as a service to the pilot project’s “actor network” in particular the sponsor BMFT; as an accompanying evaluation and as an ex-ante analysis for potential subsequent measures. During the monitoring, the evaluation team had to coordinate and process a variety of tasks, partly conflicting between implementation and evaluation. Over time, the intensity of evaluation activities increased. Targeted analyses of the equity capital supply were initiated early on, based on findings from the accompanying research. On this basis, operational ex-ante investigations were carried out with a view to designing a subsequent pilot project “Participation Capital for Young Technology Companies” (BJTU). Overall, the accompanying research and evaluation of the TOU pilot project was a comparatively time-consuming and complex process, the experimental character which included intensive investigations of the funding clientele over an extensive period of time and active participation in the funding measure, as well as the difficulty of maintaining a distanced observer status within this process. This tension is probably unavoidable in the actor-oriented development of complex new funding models, as a kind of mediator between governmental agencies, beneficiaries and industry.

## 2.5 Evaluation Community

The 1980s saw an upsurge of professionalisation and community building in the field of evaluation experiences and competencies. Most prominently, in 1986 the *American Evaluation Association (AEA)* was established, a merger of the former Evaluation Research Society and the Evaluation Network (Kingsbury 1986).

In the field of science and research evaluation the Institute for Science and Technology Research (*Institut für Wissenschafts- und Technikforschung*) at Bielefeld University has been a network hub for evaluative science studies since the mid-1970s.

In the course of the 1980s other institutes, like the Centre for Science and Technology Studies (*CWTS*) at Leiden University, helped to boost this growing field of competence and expertise.

In the field of innovation policy studies, the *Six Countries Programme (6CP<sup>2</sup>)*, established in 1974 became for many years a network hub, which provided the opportunity for professional exchange about related policy evaluations. “Within the broad area of innovation policy, the Six Countries Programme is an example for a non-issue-specific Forum with emphasis on an open-minded exchange of information, reflection and the identification of new issues/assessment of new developments” (Edler et al. 2006), including the question “how innovation policies have to be designed to create effective out-comes of such public policy”.

With financial seed support from the European Union, the University of Twente in the Netherlands established an annual *R&D Evaluation Course* in the late 1980s, to foster European and international capability and capacity-building in the field of professional research evaluation (with other European evaluation institutes as partners); the course ran until 2022.

### **3 The 1990s: Complex Multi-actor/Multi-agency Programmes and Systems-Oriented Evaluations with Formative Claims**

#### ***3.1 Political Ambitions and Developments***

In retrospect, the decade of the 1990s was characterised by a high willingness to develop and implement innovative R&I measures and a generally critical reflection on the strengths and weaknesses of national innovation systems. From a system’s perspective, the question of cooperation between the different subsystems has played a particularly important role. Accordingly, R&I policy had increasingly launched programmes since the mid-1990s that were intended to bring about structural change in science and industry through the formation of innovation networks. Partnerships of a wide variety of actors (research institutions, universities, industrial enterprises, banks, technology centres, further education institutions, etc.) aimed to effectively re-shape local, regional or sectoral innovation systems with a whole bundle of coordinated measures. Such multi-actor and/or multi-measure programmes covered either regions and/or specific missions and/or selected technologies and sectors. Unlike simple cooperation projects, such as those that have been funded since the mid-1980s in national programmes of collaborative research, complex network programmes intended to achieve a self-sustaining dynamic of the selected clusters with sustainable economic effects in global competition through start-up financing. Competence centres and cluster initiatives represent outstanding

---

<sup>2</sup>The core group of countries were Germany, Austria, Finland, the Netherlands, Sweden and the UK.

examples of these new funding instruments. One of the very challenging key funding prerequisite in those programmes was a network-like cooperation of the actors involved in the cluster that goes beyond concrete individual projects (OECD Proceedings 1999; Koschatzky 1996, 2000).

Further, since the 1990s a reorientation in the relationship between the state and publicly funded research organisations has taken place in numerous (Western) European countries, which can be summarised by the catchword New Public Management (NPM). In essence, this meant that the state reduced its (operational) control over public research institutions and granted them more autonomy in their decisions, but in return expected performance to increase (Ferlie et al. 1996). In the course of these changes in governance between scientific institutions and the state, various mechanisms were introduced, like global budgeting, performance-oriented resource allocation and goal agreements. The introduction of NPM also led to an increasing relevance of competition and performance-oriented funding, either as part of institutional funding or through increasingly important competitive third-party funds (Geuna 2001; Hicks 2012; Jongbloed and Lepori 2015; Jongbloed and Vossensteyn 2001).

### ***3.2 Main Aims and Topics of Evaluation***

In the 1990s, evaluation research and practice gained considerable momentum, driven by innovative policy approaches, on the one hand, and the increased need to legitimise public funding, especially of research and research institutions, on the other. The latter applied both in Germany and at the European level.

In Germany in the early 1990s a first meta-analysis of extant R&I policy evaluation studies showed that the number of such exercises was rising while the methodological quality of some studies and their use in policy practice remained limited (Kuhlmann et al. 1995). At the same time, however, now more often systematic so-called accompanying or monitoring evaluations were launched. They were meant to serve as a learning instrument for the continuous improvement of innovative funding approaches like EXIST, a federal programme to promote start-ups based on scientific research (Kulicke et al. 2000, 2002), Competence Centres for Nanotechnology (Bührer et al. 2000, 2001) or the Interdisciplinary Centres for Clinical Research at University Hospitals (IZKF) (Braun et al. 1997; Bührer et al. 2004).

As mentioned above, the performance of research institutions became an important topic in R&I policy and thus evaluation also turned increasingly to individual science institutions. One of the most important examples is the UK approach, today the Research Excellence Framework (REF), formerly the Research Assessment Exercise (RAE), which started as early as the 1980s, closely linked to the development of sophisticated bibliometric and other quantitative indicators (see also below). In Germany, performance of research organisations has been studied, on the one hand, with regard to the missions of research institutions (Kuhlmann and Holland

1995) and, on the other, with regard to the interaction of different research organisations within the national innovation system in the context of the so-called system evaluations (Evaluierungskommission 1998; Internationale Kommission 1999; Wissenschaftsrat 2000, 2001). Furthermore, in 1990, shortly after the German reunification, the Science Council (Wissenschaftsrat) was tasked by the government to develop recommendations for the *reorganisation of more than 130 East German non-university research institutions* (institutes of the former GDR Academy of Sciences). The aim of the recommendations was “to create as soon as possible new structures to facilitate internationally competitive basic and applied research”. (Krull 1992). The Science Council had formed an evaluation committee and discipline-oriented working groups. Their recommendations for the reorganisation of East German non-university research institutions had far-reaching consequences for the development of the innovation capacity in East Germany in the 1990s (Holland and Kuhlmann 1995).

At the European level, the 1990s showed a significant dynamic not only in R&I policy, but also in evaluations and impact analysis that covered not only the European Framework Programmes (see, for example, Georghiou and Meyer-Krahmer 1992; Bach and Lambert 1992; Stern 1993; Georghiou 1995a, b; Fayl et al. 1998, 1999; Airaghi et al. 1999; Guy and Polt 1999) but also initiatives such as COST (Cunningham et al. 1997; Cunningham and Nedeva 1999) and EUREKA (Airaghi et al. 1995; Bobe et al. 1999; Dale and Barker 1994; Dekker et al. 1991; Ormala 1993). Furthermore, various impact studies were conducted on the benefits of participating in the European Framework Programmes (for example, Larédo and Callon 1990; Georghiou et al. 1993; Reger and Kuhlman 1995; Reger et al. 1998; Luukkonen and Niskanen 1998). A special feature of the evaluation system developed within the framework of EUREKA, a European cooperation scheme formally outside the constitutional framework of the European Community, was that it systematically surveyed long-term impacts by asking the beneficiaries to answer a questionnaire not only shortly after the completion of their project, but also 3 and 5 years after the end of the project. These market impact reports were implemented in recognition of the fact that (economic) impacts often manifest themselves only a long time after the funding (Georghiou 1999).

The European Commission also supported a couple of research projects that aimed to improve the European evaluation systems. For example, the TSER-Network “Advanced Sciences for Technology Policy Planning” (ASTPP, 1997–1999), that was led by Fraunhofer ISI.

In the academic context, this decade was characterised by intensive exchange and comparative work on the respective national evaluation cultures, which in turn found their way into the further development of the respective national R&I policies (see, for example, Da Silva and Henriques 1995; Hills and Dale 1995; Kastrinos and Katsoulacos 1995; Kuhlmann 1995; Larédo and Mustar 1995; Rip and van der Meulen 1995; Sanz-Menéndez 1995; Silvani and Sirilli 1995; Steiner and Sturn 1995; van Steen and Eijffinger 1998). This work has led to the transparency of national patterns in R&I evaluation on the one hand, but also to stimulating mutual learning and the emergence of an evaluation community on the other.

### 3.3 *Applied Methods and Concepts*

Overall, evaluation research in the 1990s benefited from a scientification of evaluation practice across policy fields. American scholars were considered to be leading in this respect (Guba and Lincoln 1989; Patton 1997; Scriven 1991). In the German-speaking countries, however, considerable efforts were also made in this decade to establish a scientific basis for evaluation practice (e.g. Stockmann 2000; Widmer 1996; Wollmann 1998).

One of the new concepts that shaped this decade was the understanding of evaluation as a learning medium for the actors involved, which focused on the formative aspects of evaluation (Kuhlmann 1998a, b). The key concept of the new, expanded understanding of evaluation is “negotiation” in actor arenas.<sup>3</sup> The result of evaluations designed accordingly is no longer “a set of conclusions, recommendations, or value judgements, but rather an agenda for negotiation of those claims, concerns, and issues that have not been resolved in the hermeneutic dialectic exchanges” (Guba and Lincoln 1989, 13). Decisions are rather made as an on-going process in which competing actors interactively reach a consensus, or not, and evaluation results are one piece of information among many. Here the evaluation process, more precisely the communications of the involved actors, come to the fore and the process is deliberately designed to be “participatory” (“Participatory Evaluation”) (Patton 1997; Worthen et al. 1997). Such evaluation concepts aim above all to re-frame (Schön and Rein 1994) the orientations of corporatist and political-administrative actors. In the context of the research and innovation system, they can act as an “intelligent” contributor to the negotiation and as a coping strategy of the responsible political actors and the interested public. “Intelligent” policy development processes in this sense can furthermore be enriched by a combination with Technology Foresight and Technology Assessment (Kuhlmann et al. 1999, but also Martin 1996; Meyer-Krahmer and Reiss 1992). A concrete practical example of the combination of an (institutional) evaluation and foresight activities was the use of the results of a large-scale Delphi survey for the system evaluation of the Fraunhofer Society (Bührer 2001).

In line with the policy trends described at the beginning of this section (measures to improve the relationships between various actors within national innovation systems and a stronger performance orientation of research institutions), specific data collection and analytical methods were (further) developed in the 1990s, namely network analyses and STI (Science, Technology & Innovation indicators).

Various articles (Bührer 2002; Cabo and Bijmolt 1992; Noyons and van Raan 1996; Removille and Clarysse 1999) reflected the growing importance of network analysis at the time and this has become a standard approach in many evaluation studies. One German example, the evaluation of the interdisciplinary centres for clinical research at University hospitals (Bührer and Peter 1999), analysed in detail the cooperation behaviour of the funded individuals. This example illustrates well

---

<sup>3</sup>The following paragraph is mostly a translation of Bührer and Kuhlmann (1999), page 387 ff.

the scope and ambition of network analysis in evaluation. The aim of the survey was to examine the extent to which the implementation of the funding measure has (already) succeeded in initiating and establishing new, interdisciplinary collaborations. Among other things, the aim was to elicit possible consolidation zones within the professional relationship networks of the scientific staff active in the respective centres and to identify the central actors. In addition to describing the formal aspects of the individual networks (size, density/closeness, transitivity, number of redundant contacts, centrality, cliques, homogeneity measures), the aim was also to describe the “quality” of the individual relationships. The quality of a relationship includes the duration of the relationship, the origin of the relationship, the frequency of interaction, the (perceived) strength of the relationship, the role relationship and the transactional content.

As regards STI indicators, numerous overviews were published in the 1990s, building upon the ground-breaking work that was done during the 1980s, as described above (Brisolla 1996; Gabolde 1998; Grupp et al. 1995; Tomizawa and Niwa 1996; van Raan 1988; van Steen 1995; Schmoch et al. 1991; Schmoch 1999). In the field of bibliometrics, specific indicators on (co-)citations and inter- or multi-disciplinary collaboration have been further developed (van Raan 1993; Zulueta and Bordons 1999). The 1990s also saw the beginnings of comprehensive STI indicator reports at the European level (European Commission 1997). However, there is also a growing trend in the 1990s to critically reflect the use of quantitative indicators (for example, Pavitt 1998; Hicks 1991; Luukkonen 1991).

### 3.4 Evaluation Community

The 1990s were decisive in building an R&I evaluation community. In 1992, the *European Evaluation Society (EES)* was initiated and registered in 1996 as a non-profit association, and soon other associations followed, such as the *German Evaluation Society (DeGEval)* in 1997, with the significant participation of Stefan Kuhlmann (Fraunhofer ISI). The main “raison d’être” of the DeGEval is the promotion of evaluation research and practice by pursuing three main goals: professionalisation of evaluation, bringing together different perspectives and promoting information and exchange. The DeGEval is subdivided into individual working groups that are dedicated to discussing specific evaluation challenges in selected areas, including the R&I Policy Working Group that has existed from the very beginning.

At the same time, also specialised associations were set up, such as the *Austrian Platform Research & Technology Policy Evaluation* in 1996. This was initially an informal cooperation, deliberating on approaches and methods of evaluation and discussing evaluation practice on national and international levels. Meanwhile, the Austrian fteval platform has become a pan-European resource for evaluators and their clients, offering a variety of important services such as the systematic collection of evaluation reports, the publication of the fteval journal, which provides

timely reports on relevant evaluations and, most importantly, the organisation of major international evaluation conferences (see also section on 2010s below).

A further important milestone at the European level was the constitution of the “European RTD Evaluation Network” in 1997 that aims to promote regular exchange and cooperation between evaluation actors from the EU Member States and the Commission services. This network<sup>4</sup> is still operating.

At least two large evaluation conferences should also be mentioned as they offered the platforms for an increasing exchange on the challenges and potential solutions of R&I evaluations: the OECD organised a discourse-shaping conference on “Policy Evaluation Practices in Innovation and Technology” in 1997 (see OECD 1997) and Fraunhofer ISI organised, on behalf of the European Commission and the German Federal Ministry of Education and Science, a European conference on “Science and Technology in the New Europe” in Berlin on 7–8th June 1999 (Bührer and Kuhlmann 1999).

## **4 The 2000s: Further Development of Methods and Concepts for Capturing Complexity**

### ***4.1 Political Ambitions and Developments***

This decade in some ways was a transitional decade. R&I policies further differentiated (Polt et al. 2021) as the analysis of systems failures became more sophisticated and more and more schemes for improved, more tailored cooperation and networking were being developed. This system building approach was a further expression of the overall paradigm of national competitiveness through constantly improving the science and innovation systems which was still dominant at the beginning of the decade. At the European level, the Lisbon agenda and the Barcelona target (3% of GDP to be invested for research in Europe) were the most obvious expression of the ambition to mobilise science and innovation for the international competition of systems. Accordingly, the EU and its member States were increasingly ranked by their innovation activities and capabilities, most notably demonstrated by the establishment of the European Innovation Scoreboard in 2001. Though this was not a tool to evaluate policy, it signified the focus of R&I policy and its supporting evaluation across Europe. Only once the Lisbon agenda did not deliver on its claim, moves were made towards an orientation of science and innovation policy towards “Grand Challenges” (European Commission and Directorate-General for Research and Innovation 2006). Slowly, but accelerated by the fallout of the financial crisis in 2008, European and national R&I policies moved towards a policy mode that linked

---

<sup>4</sup>EUEvalnet, [https://research-and-innovation.ec.europa.eu/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/evaluation-impact-assessment-and-monitoring/evaluation-network\\_en](https://research-and-innovation.ec.europa.eu/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/evaluation-impact-assessment-and-monitoring/evaluation-network_en)

support for research and innovation to societal goals (see the next section) and thus linked policies which were system enhancing with those that were system directing (see below).

## 4.2 *Main Aims and Topics of Evaluation*

A first major development was to continue and further develop the formative and summative evaluation of multi-actor/multi-measure systems. This included cluster evaluations, the evaluation of competence centres (Bühler et al. 2002), approaches to knowledge and technology transfer (Kulicke 2002) and the comprehensive evaluation of the Austrian programmes Kplus which at the time were advanced multi-actor programmes (Edler et al. 2004; Edler and Lo 2004; Biegelbauer 2007). Given the complexity of those multi-actor, multi-measure programmes, policy makers and evaluators co-generated principles of design, implementation and evaluation<sup>5</sup> (Bühler et al. 2004). Although the evaluations had a clear summative mandate, the formative dimension gained considerable weight, as evaluation became part of creating mutual understanding and expectation management at the level of individual projects, but much more so at the level of joint system building.

In this decade, the system level became increasingly important for evaluators and R&I policy analysts. More and more, systematic, robust and valid evaluation at the systems level was called for (Arnold 2004; Feller 2007). At the end of the previous decade, the evaluation of the Norwegian Research Council was, given the broad remit of the council, a comprehensive systems evaluation (Arnold et al. 2001; Kuhlmann and Arnold 2001). One decade later, the evaluation of the Austrian science and innovation system was, and still is as of today, the most comprehensive and holistic evaluation of a science and innovation system, its governance structures and instrumentation.<sup>6</sup> Similar developments were to be witnessed in the USA, with attempts to focus on the multi-level interaction of policy interventions and the overall systems level additionality (see the special issue by Rogers and Jordan 2010). However, those in-depth systematic system-level analyses remained the exception rather than the rule at the national level.

In line with the policy focus on the systems level, in the early 2000s, both the EU and the OECD developed and financed benchmarking approaches to assess and compare how policies influence the innovation system (Molas-Gallart and Davies 2006, pp 73–74). In this approach, rather complex relations between input and output within innovation systems are operationalised, standardised and compared between countries. The most basic, but most continuous example is the aforementioned Innovation Scoreboard. Another prominent example is the benchmarking of

---

<sup>5</sup>Edler et al. (2004) provide an overview of many of those measures and develop design and evaluation principles.

<sup>6</sup>For a summary of major aspects, see the Newsletter of the Austrian Platform for Research and Technology Evaluation, 34, 2010. <https://repository.fteval.at/72/%20>

industry-science relationships that is based on theoretical assumptions of the interaction between science and industry and a set of indicators developed to map these relationships. While those benchmarking exercises have limited analytical value, they have strongly influenced the R&I policy discourse across the OECD world (OECD 2002).

To further support system-level evaluation and steering, country benchmarking and national and regional review exercises were intended to provide learning opportunities for policy makers. At the European level, in the 2000s the so-called “policy mix” reviews performed a similar learning exercise in a number of countries “on demand”, a light-touch approach based on international peer reviews. Those European country reviews did not amount to the expert led application of sophisticated methods, rather they were expert supported peer-review exercises underpinned by a number of background analytical reports most often provided by the country reviewed. Also in the 2000s, the OECD launched their “innovation policy reviews” which claim to be a “comprehensive assessment of the innovation system of individual OECD member and partner countries, focusing on the role of government. [...] Each review identifies good practices from which other countries can learn”.<sup>7</sup> Those OECD reports, while also underpinned by self-assessment and background material of the countries under review, were—and still are—led by R&I policy evaluation experts.

Below the level of entire R&I systems, but in fact as a building block for systems evaluation, a line of work intensified that tried to evaluate the interplay of diverse interventions. A special issue of *Research Evaluation* highlighted and supported the claim to analyse systematically the interplay of instruments (Vonortas et al. 2007). In particular in the USA there were a few attempts to mobilise evaluation techniques for portfolio management (Bozeman and Rogers 2001) and a number of quantitative portfolio evaluations were conducted for public research organisations (Ruegg 2006, 2007; Hage et al. 2007), with the claim to mobilise portfolio evaluation to better understand system level impact (Wagner and Jordan 2009).

At the beginning of the decade, evaluators and policy makers intensified efforts to understand wider societal impacts of publicly funded research. Rather than developing further, more sophisticated quantitative analyses, more qualitative, indirect approaches were established. A body of literature emerged that focused on tracing the processes supporting longer-term change and analysing the conditions influencing and interacting with the initial research process and the use of research results. The most prominent effort in this regard was the SIAMPI project funded by the European Commission (Cozzens 2010; Spaapen et al. 2011; Spaapen and van Drooge 2011), which focused on productive interactions between science and non-scientific stakeholders as the key driving force and a necessary precondition for the effective use of research results within the science system but also in other policy sectors. The term “productive” characterises the interactions and is only seen as successful once the research results (including knowledge) have been actively used and

---

<sup>7</sup> See <https://www.oecd.org/sti/inno/oecd-reviews-of-innovation-policy.htm>

have induced a change in behaviour. This approach was subsequently applied to a range of examples (Molas-Gallart and Tang 2011).

However, many developments outlined for the previous decade continued and started to create tensions in the R&I Policy community. On the one hand, as indicated above, there were an increasing number of evaluations that tried to capture dynamics at the systems level and understand the role of policy and policy portfolios for system developments. This was also linked to an intensified effort to support policy making throughout the entire cycle with in-depth formative approaches. On the other hand, the policy-making system, with notable exceptions, continued to demand a summative assessment, a quantitative number, to justify and legitimise R&I policy spending (Molas-Gallart and Davies 2006; Edler et al. 2010). Further, within the R&I policy evaluation community, calls for more robust quantitative and experimental methods emerged, asking for randomised controlled trials also in R&I policy evaluation. Especially as the evaluation practice had gained momentum and diversified in the previous years, the diverse claims of R&I policy evaluation and the growing tension between those claims led to an increased self-reflection in the community, with key collective events (2006 in Austria) and stock-taking and analysis exercises (Edler et al. 2010; Polt 2003).

### ***4.3 Applied Methods and Concepts***

In the field of bibliometrics, many studies began to integrate the individual level by studying mobility patterns based on CV data. This was not only an innovation in terms of the database and methodology but also the level of observation (individuals instead of institutions) (Cañibano and Bozeman 2009; Lepori and Probst 2009; Sandström 2009; Zellner 2002).

The INNO-Appraisal analysis of evaluations in innovation policy conducted between 2002 and 2007 found that most evaluations focused on economic and sometimes technological impact, societal impact, in contrast, was not very common and if so was concerned with the number of jobs created. Further societal or even transformational issues were still the exception (Edler et al. 2010). The analysis further demonstrated a tension in the R&I policy evaluation practice just outlined. While there was an increasing claim that evaluation is supposed to be formative and take into account the complex system conditions, the INNO-Appraisal study found that summative evaluations using quantitative methods, thus delivering a quantitative result, a “number” for policy makers, were more likely to be used in the policy-making process. This is in line with the analysis of Molas-Gallart and Davies (2006), who also cite two examples of evaluation in the early 2000s in which policy makers as clients disregarded the formative and system sensitive approach of the evaluation and asked for a summative, ideally quantitative assessment to be used in the political process. Thus in the 2000s a tension intensified between the claim and ambition of evaluation experts rooted in the innovation systems paradigm and policy makers

in need of quantified, ideally unambiguous and short-term impacts of their interventions.

#### **4.4 Evaluation Community**

This decade was characterised by a growing interest in the scope and effects of evaluation practices themselves, both within the evaluation community (Edler et al. 2008), in Europe and North America (Shapira and Kuhlmann 2003), and within the community of policy makers who commissioned evaluation studies (Edler et al. 2010). The diversification of evaluation claims and evaluation methods, in particular but not exclusively at the European level, led to a number of projects taking stock of evaluation practices, making those diverse approaches widely accessible to the community, assessing the appropriateness of approaches and suggesting further adjustments to meet the needs of a changing policy environment. At the beginning of the 2000s, two European funded projects analysed the evaluation practices in Europe (Polt 2003) at the European and the national level and provided access to the breadth of evaluation methods (Fahrenkrog et al. 2002, EPuB<sup>8</sup>) as well as setting out future developments to bring evaluation practice in line with R&I policy developments across Europe (Georghiou et al. 2002, ASIF<sup>9</sup>). ASIF demonstrated that the policy and evaluation rationales were now firmly grounded in the innovation systems and “failure” paradigm. It also showed that the ambition of evaluation was increasingly to determine the net effect of interventions and to capture the broader economic and in particular societal effects of R&I policy (Polt 2003, also Rojo 2003). The EpUB study led to a “toolbox” for R&I policy evaluation based on European experiences, while at the same time 10 years of evaluation of the US American Advanced Technology Programme had led to a “toolkit” (Ruegg and Feller 2003). Some years later, the European Commission ordered the aforementioned study INNO-Appraisal to systematically take stock of R&I policy evaluation practices across European countries and at the European level (Edler et al. 2010).

The significance of those exercises is twofold. First, the R&I policy community had reached a state of maturity to come forward with major, almost standardised “toolkits” and recommendations as to when to use what approaches (Miles et al. 2006; Molas-Gallart and Davies 2006). Second, and maybe even more importantly, it signifies the self-awareness of the R&I community in Europe and the USA as an epistemic community, linked together by a convergent understanding of theoretical justifications for R&I policy interventions and a set of techniques to be shared and applied. This institutionalisation of the community was further supported by major

---

<sup>8</sup>EPuB: Socio-Economic Evaluation of Public Research, Technology and Development (RTD) Policies

<sup>9</sup>ASIF: Assessing the Socio-Economic Impact of the Framework Programme

international conferences initiated by the Austrian R&I policy evaluation community (2003 and 2006).

## **5 From the 2010s Until Today: Is Accountability Back? Evaluation Between Legitimacy and Learning**

### ***5.1 Political Ambitions and Developments***

The years from 2010 onwards were characterised by a growing discussion on R&I policy as an instrument that should bring forth solutions for societal grand challenges such as climate change, demographic change or human diseases. Concepts such as “mission-orientation innovation policies” (Mazzucato 2018) and “transformative innovation policies” (Diercks et al. 2019) shifted the sphere of influence of R&I policies beyond the boundaries of the science system. Hence, research results were expected to influence non-research performing stakeholders and to have effects on non-research related sectors. As a consequence, demonstrating the impacts of research and innovation activities became increasingly important for R&I stakeholders (see discussion in the chapter by Lindner et al. 2024 in this anthology).

In parallel a, not entirely disconnected, discussion emerged, questioning the appropriateness of existing research assessment approaches and metrics (“the metric tide” (Wilsdon et al. 2015)) for assessing scientific output almost solely based on peer-reviewed journal publications. The unease towards the research assessment mechanisms (discussed already in the previous sections of this chapter) was reflected in two declarations stemming from the research community itself: The San Francisco Declaration on Research Assessment (DORA 2012) and the Leiden Manifesto for research metrics (2015)

In order to address the changes in the R&I policy landscape and methodological implications of the shift towards impact (namely the attribution challenge), the evaluation community directed its attention to developing and refining conceptual frameworks suitable to analyse the processes leading to longer-term impacts of research. It had to develop metrics capable of measuring those longer-term effects. Furthermore, the debate on the function of programme and policy evaluation and the roles evaluators opened up again in the discussions related to sustainability transition and responsible research and innovation (RRI).

### ***5.2 Main Aims and Topics of Evaluation***

As mentioned in the preceding chapter, societal impacts of funded research became a predominant topic at the turn of the decade. The observation that a longer-term impact of research had not been adequately addressed in evaluation studies of the

early 2010s (Arnold 2012) is a result of several challenges which confront impact measurement of research:

First, there was no commonly shared definition of what impact of research means, the dividing lines reflecting the researchers' disciplinary backgrounds (Arnold 2011; Bornmann 2013; Donovan 2011; Flecha et al. 2018). In recent evaluation practice two main aspects have been emphasised with regard to impact, as e.g. put forward in the Glossary of the Evaluation Standards Kohlweg (2019): A) an impact can be positive or negative, primary or secondary, intended or not by the funded intervention. B) It is an effect that materialises on stakeholders not involved in the research processes and reaching out beyond the science system itself.<sup>10</sup> Building on earlier work mentioned above, the term "societal impact"<sup>11</sup> further grew in popularity in the last years, however, no commonly agreed definition exists for it until today (Bornmann 2013; Bühner et al. 2022a, b; Joly and Matt 2017; Muhonen et al. 2019; Reale et al. 2018).

A second challenge related to the attribution problems is the difficulty to establish a causal relation between a particular impulse such as a research result or a funding programme and a longer (societal) change. This was not a new discussion in the evaluation literature, but became more critical when focusing on longer-term outcomes and impacts of R&I funding. The consecutive challenges related to the evaluation of mission and challenge-oriented interventions have been raised by several authors lately: expected effects in various policy sectors (multiple impact dimensions (Bühner et al. 2022a, b); long-time horizons which effects would materialise that go beyond the evaluation exercise; changes that are beyond the control of the stakeholder initially involved in the research and innovation processes; interference and interaction of the research with other funding or sectoral policies (Amanatidou et al. 2014; Arnold 2018; Donovan 2011; Donovan and Hanney 2011; JIIP 2018; Molas-Gallart et al. 2021; Reale et al. 2014; Wittmann et al. 2022).

In consequence, the focus shifted away from traditional approaches measuring economic effects through well-established STI indicators. Rather, the concept of contribution as opposed to attribution has gained importance in the evaluation literature and the methodologies developed allow to analyse the contribution that research makes on broader socio-economic societal changes.

In line with the more ambitious attempts to capture impact, a concept that has grown in popularity in R&I programme evaluation in the last 10 years is one of the impact pathway modelling. Building on a "theory of change" (of the funded intervention) (Chen and Rossi 1983; Weiss 1993, 1997) idealised pathways are described showing how a stimulus in form of research results can contribute to longer-term changes. The pathways rely on a set of assumptions how effects can manifest

---

<sup>10</sup> fteval Evaluation Standards 2019; Glossary: "Positive and negative, primary and secondary long-term effects resulting directly or indirectly, intentionally or unintentionally from an intervention. OR effects beyond the intended reach of the evaluation object, i.e. on things (systems, institutions, organisations, individuals, outputs etc.) where the evaluated object has been a contributing factor".

<sup>11</sup> Also the term of "social return" is used (Donovan 2011) or "socio-economic impact" Reale et al. (2018).

themselves. Its particular strength lies in an explicit consideration of external conditions interfering with the initial research process and subsequent diffusion of research results. The focus of these approaches is on understanding the underlying mechanisms and processes leading (or not) to societal changes. Recent work of Belcher et al. (2020) extended the pathway logic by drawing attention to the stakeholders involved in the evaluated intervention and introduced the reasoning in spheres of dwindling control when moving up the impact pathway. The theory-based evaluation approaches thereby contrast the standard evaluation approaches which rely on experimental or quasi-experimental methods and aim at quantifying effects on the macro-economic level. However, they are not interested in how and why the measured effects develop (“black-box approach”) (Chen and Rossi 1983).

An attempt to put into practice theory-based-evaluations analysing the changes along the theory of change is the method of contribution analysis (CA) (Mayne 2012). Studies applying this concept to R&I interventions remain scarce and raise doubts whether the idea of building a contrafactual situation relying on the theory of change can be used to measure contributions in complex systems such as the R&I field (Morton 2015). The effective use of CA in evaluation studies is hampered by the enormous costs that a rigorous analysis of processes implies (Delahais and Toulemonde 2012).

With the normative turn in innovation policy (Borrás and Edler 2020; Schot and Steinmueller 2018), whereby research and innovation policy is mobilised to support specific societal missions, most prominently in the field of sustainability and climate change mitigation, there has been again a growing interest in comprehensive formative evaluation frameworks (Amanatidou et al. 2014; Arnold 2018; Dinges et al. 2020a; JIIP 2018; Magro and Wilson 2013; Mickwitz et al. 2021; Molas-Gallart et al. 2021; Patton 2019; Wittmann et al. 2022). The focus is on the continuous improvement of the design and the implementation of an intervention. Capacity-building and learning of stakeholders involved in the evaluation becomes more important (Patton 2011). This impacts also the role of the evaluators themselves who become mediators with different interests but also co-producers in the evaluation process (Guba and Lincoln 1989). In addition, ex-ante impact assessments are receiving more attention using impact pathways to conceptualise possible future effects (JIIP 2018; Weber and Polt 2014; Wittmann et al. 2022).

### ***5.3 Applied Methods and Concepts***

With the shifts in R&I policy just described, a growing number of evaluations until today apply concepts of theory-based approaches and impact pathways to different research and innovation programme funding. Increasingly, studies across different fields and categories of policy measures apply the aforementioned approach of productive interaction (Molas-Gallart and Tang 2011; Kroll et al. 2022). Others further develop the impact pathways to their specific evaluation domain or combine both to evaluate process factors and measure impact at the same time (Muhonen et al. 2019). The mainstreaming of impact pathways reasoning for the evaluation in the

R&I field can be seen in the development of an assessment framework for the latest EU Research Funding Framework Programme, Horizon Europe, building “Key Impact Pathways” (Bruno and Kadunc 2019).<sup>12</sup> Further, the ASIRPA project has been a seminal project that used the impact pathways concept to trace the impacts of a research performing organisation by using case study methodology and cross-cases comparison (Joly et al. 2015). Indeed, the legitimacy of qualitative approaches as a method for impact evaluation studies is extended in the 2010s (Warta and Philipp 2014), a prominent example being the use of impact case studies and narrative in the Research Excellence Framework (REF) (Pinar and Horne 2022; Wilsdon et al. 2015). Nevertheless, it should be emphasised that aggregated econometric top-down models are also envisaged, at least for measuring the EU Key Impact Pathways (see, for example, the Nemesis model<sup>13</sup>).

To capture the increasing transformative ambition and pathway complexity of R&I policy intervention, evaluation practice more and more applies a mixed-method approach and uses triangulated data and analysis methods to capture societal impacts and transformation through R&I policies (Arnold 2018; Joly and Matt 2017; Magro and Wilson 2013; Seus and Bühler 2021; Warta and Philipp 2014).

The mid-2010s saw also a lively discussion on metrics used to measure research impact and the limitation of the traditionally used indicators (Hicks et al. 2015; Wilsdon et al. 2015). Although indicators measuring output and quality of research are, for the obvious reasons discussed in this chapter, only one type of assessment criteria used in evaluation studies, they nonetheless remain a fundament of most evaluations dealing with research funding. For example, the exploitation of bibliometric or patent databases measuring publication outputs and quality and cooperating patterns are continuously used in evaluation studies and the related methods of analysis are further fine-tuned. Today it is a standard tool in research funding evaluations (see, for example, the Evaluation of the Erwin Schrödinger Fellowship (Meyer and Bühler 2014)).

It is important to note that new metric avenues open particularly with regard to new data sources that can be summarised under the keyword “Big Data”. This includes altmetrics or more general sources from the web or the use of quantitative text analysis methods to analyse big sets of documents. Examples of studies can be found in the SIPER repository (e.g. Bühler et al. 2021 and Feidenheimer et al. 2018). Furthermore, the combination of different data sets offer new possibilities for analysis. One example is the evaluation of the START-Wittgenstein programme (Seus et al. 2016) that used the bibliometric database Scopus to create a control group to be compared with the funded researchers. This was a methodological novelty, as evaluation studies usually only make use (if at all) of comparison groups drawing most of the time on non-successful applicants.

The last decade also saw a continued debate on the value of Randomised Controlled Trials (RCTs) in R&I policy evaluations. Those approaches are

---

<sup>12</sup>It must be noted that the programme theory approach is not new to the evaluation approaches of the EU, but a core approach in the European Structural Funds Evaluation and since 2017 also a standard component of the “EU better regulation guidelines”.

<sup>13</sup><https://web.jrc.ec.europa.eu/policy-model-inventory/explore/models/model-nemesis>

experimental in character which imply the inclusion of the RCT design right from the beginning of the intervention and, accordingly, have to be planned by the funding organisations before the launch of a programme (frontload of the evaluators work). Further, those approaches need a large number of funded entities and the possibility to experiment with randomised funding in the first place. The best known initiative promoting experimental innovation policy is the RCT Innovation Growth Lab (IGL) at Nesta's, which was founded in 2014 and its IGL Trails Database collects examples from experimentation and controlled trials.<sup>14</sup> The Austrian Research Promotion Agency (FFG) has engaged itself in this journey for three of its programmes (Landon and Hochreiter 2022). However, as pointed out by Firpo and Phipps (2019) RCTs have not been the main focus of R&I policy and methodological developments are rare with a few exceptions, such as the RCT+ approach (Bakhshi et al. 2015). Whether it will become an interesting method to fund and consecutively measure the impact of mission and transformation-oriented policies will be seen in the coming years. It will, in any case, be limited to very specific research questions and a very specific form of funding programmes, with a large number of funded and non-funded entities and the political opportunity for the experiments to be executed.

With the increasing importance of formative evaluation and ex-ante impact assessment, the demand has grown for broader data collection and analysis methods used in evaluation, and especially to include foresight methods, i.e. scenario building or Delphi surveys (Dinges et al. 2020b; JIIP 2018; Patton 2019; Weber and Polt 2014). These authors emphasised the fact that foresight methods can enrich the evaluation as they help explore the interaction between the different interventions and policies (policy mix) and acknowledge the complexity of longer-term change. Furthermore, inclusiveness, stakeholder involvement (opening up to non-traditional R&I stakeholders) and participation in evaluation design and implementation (co-creation) gained in importance as well as the evaluation of experiments and trans-disciplinary research processes (Belcher and Hughes 2021).

Finally, this decade also saw a persistent interest in collating and learning from existing evaluation studies. One prominent example is the Handbook of Policy Impact (Edler et al. 2016), which, through a unified "evaluation synthesis approach", shows the diversity of effects generated by different policy instruments. A second example is the SIPER database ([www.si-per.eu](http://www.si-per.eu)) which allows access to a wide variety of evaluation studies to be used for different analyses in terms of evaluation practice and innovation policy effects.

## 5.4 Evaluation Community

The R&I evaluation community has grown into a mature and stable community in which the changes in the R&I policy field resonate. Looking at the fteval's mission statement of 2014, one notices the shift from a narrow technology policy focus

---

<sup>14</sup><https://innovationgrowthlab.org/igl-database-v2>

towards a broader definition of innovation (Warta and Philipp 2014). The field became further institutionalised. Most notably, between 2013 and 2022, the European R&I Policy Evaluation Conference (in 2013 still named “evaluation of STI policies, instruments and organisations”) took place four times (in 2013, 2017, 2019 and 2022) and has established itself in a biennial rhythm. It is a forum which brings together researchers, evaluation practitioners and policy makers. The evaluation standards elaborated by the fteval members were updated for the last time in 2019, reflecting the changes in the R&I landscape and taking up the main discussion described in this chapter. In 2016 the Network for Advancing & Evaluating the Societal Impact of Science (AESIS) brought together the organisations involved in the evaluation of impact, research strategy and funding of research. Although the R&I evaluation community is today a consolidated community, it has nevertheless opened up to well-established evaluation practices used in other policy domains (especially social sector evaluation and development assistance) such as the concepts of theory-of-change modelling and applied them to R&I evaluations. As can be seen in numerous contributions of the latest European R&I evaluation conference (ftval-journal 2022), the evaluation community is reacting to these developments by discussing new approaches to measure effects and provide evidence for policy learning.

## 6 Summary

The evaluation practice in R&I policy has shown a remarkable development in terms of its very purpose, the questions asked, the methods used and the institutionalisation of the community. As we have seen, throughout the last decades, this development has naturally interacted with the development in R&I policy, practice and theory.

In sum, in the 1960s, 1970s and 1980s the purposive evaluation of public research and innovation activities and policies took shape. It covered a growing scope of topics and it gradually became more professional, from policy planning to policy implementation and policy impact evaluation; from science to technology, research and innovation processes as subjects of evaluation; from academic studies (sociology of science; economics; policy studies) to systematic professional research projects, often on behalf of governmental actors such as a service, nurtured by an internationally growing community of experts.

The 1990s took up and further developed the central approaches and work from the previous decades. However, the 1990s also saw important new developments in the field of policies (e.g. multi-actor, multi-measure programmes) as well as the strengthening of a national and international evaluation community. The decade was characterised by increased reflection on evaluation concepts and methods and the subsequent methodological and conceptual developments. On the one hand, these have brought the idea of evaluation as a core element of policy intelligence into focus, but on the other hand, they have also further developed numerous quantitative and indicator-based methods.

The decade between 2000 and 2010 was one of transition. R&I policy, while further differentiating, slowly moved from a system enhancement approach, focused mainly on economic and system effects, to an approach that was increasingly interested in societal impacts. In terms of strengthening the performance and structure of the science and innovation systems, policy measures targeting complex network structures led to improved evaluation methods to capture this complexity. Increasingly attempts were made to assess and benchmark entire systems and the role of R&I policies. At the same time, attempts intensified to capture impact and impact pathways of research, which foreshadowed a more ambitious R&I policy in the years to come. The policy and evaluation community also intensified stock-taking and learning exercises to reflect the opportunities and weaknesses of the evaluation practice and to further standardise the toolkit and strengthen the self-awareness of the evaluation practitioners as an epistemic community.

In line with the increased societal ambition of R&I policy, the years from 2010 to today have been marked by an intensified discussion on how to measure broader impacts of research and R&I policies. As a consequence, the evaluation community turns its attention to developing and refining conceptual frameworks suitable for analysing the processes leading to longer-term impacts of research and to develop metrics capable of measuring those long-term effects. Towards the end of the decade, the formative function of evaluation focusing on understanding funding mechanisms and distribution processes grew in importance reflecting the difficulties in measuring impacts and the need for advice how to govern the transition towards sustainability. The evaluation community in the R&I field has grown into an epistemic community using a wide range of qualitative and quantitative methods and exploring the possibilities of big data analysis.

## **7 Reflections and Outlook**

### ***7.1 Co-evolution of Policy and Evaluation: Interaction and Interdependencies***

Our review of the emergence and development of evaluation concepts and efforts in the domain of research and innovation policy in the last 60 years suggests that, often, policy ambitions and designs, on the one hand, and evaluation methods and practices, on the other, go hand in hand; they co-evolve. Creative evaluation experts, based in academia or in consultancy, have developed mixed-method approaches to understand and model complex input/output relations of R&I policy interventions. They have advanced qualitative as well as quantitative data analysis methods, and they have explored new data sources. Vice versa, enlightened and creative policy-makers have used intelligence input from such evaluation studies to conceptualise

and implement ever more sophisticated policy instruments that aim to stimulate effective and competitive research and innovation activities.

This co-evolution, though, occurs also with dialectic frictions: diverging policy targets and designs by competing policy actors (different ministries, agencies) interact with diverging evaluation concepts and designs by different experts, with diverse epistemological backgrounds. Evaluation results can and should inform the owner of an evaluated policy, however, evaluation results will also be perceived by other policy actors. This in turn can confuse, but also stimulate debate and competition between policy approaches and evaluation designs. Ideally, evaluation exercises would adopt a “moderation” function in policy arenas with competing actors. Evaluation can function as a strategic intelligence interface in a “multiple perspective framework” (Kuhlmann 1998a). Helping policy actors with the means of evaluation studies to reflect on the perspectives of other actors in the R&I policy arena can enrich and facilitate a re-framing (Schön and Rein 1994) of the orientations of corporatist and policy-administrative actors (Kuhlmann 1998a), a precondition for policy learning (Padilla 2016).

## 7.2 *Critical Developments and Challenges*

Over the past decades, the practice of evaluating publicly funded research and innovation has increased significantly. There is no doubt that evaluation exercises are necessary and justified. Funded projects must demonstrate that the use of taxpayers’ money is efficient and meaningful. But what we also see today is that more and more evaluation exercises are done mechanistically by research or policy administrators or specialised service providers without a sound understanding of the scientific process or content, but rather following a given, more or less uniform set of technical criteria. This can lead to scientists performing to meet those criteria and stifle scientific creativity. The important search for evidence of public policy impacts can, if done mechanistically, revert and incite evaluated organisations to fabricate “policy-driven evidence” (Strassheim and Kettunen 2014).

Another worrying development in recent evaluation practice is the definition and application of a gold standard of evaluation in R&I policy which then puts pressure on the evaluation community and even on the design of programmes. More concretely, randomised controlled trials, as we have known them for decades, in health or education research policy have been increasingly asked for an innovation and entrepreneurship policy as well (Bravo-Biosca 2019; Landon and Hochreiter 2022; Roelandt and van der Wiel 2020). While those approaches have their merits for interventions in health and education, where the population is far more numerous and experiments can be on a large scale, and where large-scale intervention has an identifiable impact on individual units, their merit for science and innovation policy is limited. Nevertheless, the normative pressure to use RCTs has grown in a number

of countries (Leko et al. 2016)<sup>15</sup> which—in our observation—puts pressure on policy makers to design programmes to fit the gold standard RCT or to apply RCTs even if, given the low numbers of units involved in trials, RCTs do not really fit the programme. Here, policy risks follow evaluation fashion rather than problem adequacy.

A further challenge is that the evaluation of research organisations and their performance, such as those carried out in the United Kingdom since the 2010s, but also in the Netherlands, have far-reaching implications not only for individual scientists and their institutions, but also for the entire research landscape. Many experts argue that research assessment procedures went out of control. In the Netherlands, for example, the initiative “Science in Change” ([www.scienceintransition.nl/english](http://www.scienceintransition.nl/english)) has addressed this prominently. The main concern of the initiative was that evaluation mechanisms used in academia are too mechanistic, if the quality of science is primarily measured by highly formalised performance indicators.

Such mechanistic, quantitative approaches can, inter alia, lead to an underestimation of the different dimensions of science as well as of societal relevance. Accordingly, there has recently been an increase in discourses on “responsible metrics”, which primarily aim to make the existing procedures for evaluating science more responsible, as has been called for by numerous researchers in the context of DORA and the Leiden Manifesto, among others (Wilsdon et al. 2017; Hicks et al. 2015; DORA (San Francisco Declaration on Research Assessment 2012, (<https://sfdora.org/read>))). The current initiative at the EU level to improve research assessment is also part of this series of activities. More than 350 organisations from over 40 countries have signalled their interest in participating in this initiative and in July 2022 the final version of the agreement on the reforming research assessment was published ([https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20\\_en](https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20_en)).

### 7.3 *Future Developments*

Topics in evaluation research and practice for the years to come in particular support R&I policy in developing and implementing meaningful transformative or mission-oriented approaches. This will mean more formative work—including a link to foresight—and new attempts to understand, identify and monitor conditions for the pathways of impact. Additionally, the evaluation of action research and transdisciplinary research processes (Belcher and Hughes 2021) become important. At the same time, with the increased ambition of R&I policy to contribute to transformation and to achieve missions, there is more pressure to “deliver”. Consequently, the

---

<sup>15</sup> See, for example, the broad internationally impactful activity by the think tank NESTA, <https://www.nesta.org.uk/toolkit/running-randomised-controlled-trials-in-innovation-entrepreneurship-and-growth/>

formative work will need a range of (new) summative endeavours to deliver on “measuring” the success of policy intervention and R&I activities in terms of transformative pathways, etc. There is also a need to open up the R&I evaluation community to experts in other policy areas (such as health or environment) to show effects of R&I on transformations. Also, those experts can improve the formative work.

In this context, it is increasingly challenging and important to address the attribution problem. This is why approaches like theory-based evaluations and modelling, already described above, are gaining in importance. This is because, first, the desired impacts of transformation-oriented R&I policies are by definition long-term effects and are thus often not yet available when evaluation is asked to measure impact. Second, any given research and innovation policy intervention is often only one factor among many influencing the form and speed of the transformation. Here, new approaches and new combinations are needed. The modelling of societal-impact scenarios can draw on a well-established approach of explicating assumptions that allows for a combination of R&I policy with other important determinants of impact. The plurality of scenarios then illustrates the future openness of the long-term impacts. Such an approach brings together qualitative bottom-up information with system analyses, i.e. it also addresses qualitative factors and does not intend to cover all relevant factors with one *uno actu* top-down (econometric) model. Thus, building on the experience with system-analytical modelling to deal with the complexity of impact relationships in the context of ex-ante sustainability impact assessments (e.g. in energy and mobility scenarios) can also be a suitable approach for making societal impacts tangible.

Further, there are increasing calls in evaluation research to conduct evaluations (even more so than in the past) with the active involvement of different stakeholder groups, a trend that is being intensified with the current claims of R&I policy to support transformation and thus to engage with a broader and more complex context and actor constellation. This stakeholder involvement can consist of asking non-academic user groups about their perception of the “impact” of research. It can also include a process component, namely that stakeholder groups and/or those affected are systematically involved during the design, implementation and validation phases of an evaluation. Such procedures are certainly easier to implement in formative evaluations than in summative ones, even if they are also possible in the latter.

Finally, our necessarily abridged and somewhat selective overview of five decades of evaluation practice in R&I policy has shown that we need continuous, in-built critical reflection on the (continuously changing) role of evaluation in R&I policy and what it can, and cannot, deliver.

Given the increased ambition of R&I policy, it remains critical that the practice of evaluation shall continue to interact productively with policy practice to co-develop both evaluation and policy in meaningful ways. Thus, it is more important than ever that evaluation exercises are not seen as necessary, mechanistic services following pre-defined scripts, but rather are understood as essential processes to learn and to further develop, apply and communicate methods and concepts.

## References

- Airaghi A, Becker C, Bobe B, Georghiou L, Kemper V, Rönström T, Wiegart N (1995) EUREKA evaluation report. Brussels
- Airaghi A, Busch, Niels E, Georghiou L, Kuhlmann S, Ledoux MJ, van Raan AFJ, Baptista JV (1999) Options and limits for assessing the socio-economic impact of European RTD programmes. Report to the European Commission, DG XII, Evaluation Unit, Brussels/Luxembourg
- Amanatidou E, Cunningham P, Gök A, Garefi I (2014) Using evaluation research as a means for policy analysis in a 'new' mission-oriented policy context. *Minerva* 52:419–438. <https://doi.org/10.1007/s11024-014-9258-x>
- Arnold E (2004) Evaluating research and innovation policy: A systems world needs systems evaluations. *Res Eval* 13:3–17
- Arnold E (2011) Understanding the long-term impacts of the EU framework programme of research and technological development. Inaugural lecture. School of Management and Governance of the University of Twente, Twente
- Arnold E (2012) Understanding long-term impacts of R&D funding: The EU framework programme. *Res Eval* 21:332–343. <https://doi.org/10.1093/reseval/rvs025>
- Arnold E (2018) How should we evaluate complex programmes for innovation and sociotechnical transitions? *Technopolis*. Checked on 21 May 2022
- Arnold E, Kuhlman S, van der Meulen B (2001) A singular council: Evaluation of the Research Council of Norway. Technopolis, Brighton, England
- Bach L, Lambert G (1992) Evaluation of the economic effects of large R&D programmes: The case of the European space programme. *Res Eval* 2:17–26
- Bakhshi H, Edwards JS, Roper S, Scully J, Shaw D, Morley L, Rathbone N (2015) Assessing an experimental approach to industrial policy evaluation: Applying RCT+ to the case of Creative Credits. *Res Policy* 44:1462–1472. <https://doi.org/10.1016/j.respol.2015.04.004>
- Becher G, Kuhlmann S (eds) (1995) Evaluation of technology policy programmes in Germany. Economics of science, technology and innovation, vol 4. Springer Dordrecht, Dordrecht
- Belcher BM, Hughes K (2021) Understanding and evaluating the impact of integrated problem-oriented research programmes: Concepts and considerations. *Res Eval* 30:154–168. <https://doi.org/10.1093/reseval/rvaa024>
- Belcher BM, Davel R, Claus R (2020) A refined method for theory-based evaluation of the societal impacts of research. *MethodsX* 7:100788. <https://doi.org/10.1016/j.mex.2020.100788>
- Biegelbauer P (2007) Learning from abroad: the Austrian competence centre programme Kplus. *Sci Public Policy* 34:606–618
- Bobe B, Bühner S, Chelli R, Eriksen H, Peksen Z, Nedeva M, Sand F (1999) The continuous and systematic evaluation of EUREKA. Annual impact report 1999, Brussels
- Boekholt P (2010) The evolution of innovation paradigms and their influence on research, technological development and innovation policy instruments. In: Smits RE, Kuhlmann S, Shapira P (eds) *The theory and practice of innovation policy*. Edward Elgar, Cheltenham
- Bornmann L (2013) What is societal impact of research and how can it be assessed? a literature survey. *J Am Soc Inf Sci Tec* 64:217–233. <https://doi.org/10.1002/asi.22803>
- Borrás S, Edler J (2020) The roles of the state in the governance of socio-technical systems' transformation. *Res Policy* 49:103971. <https://doi.org/10.1016/j.respol.2020.103971>
- Bozeman B, Rogers J (2001) Strategic management of government-sponsored R&D portfolios. *Environ Plan C Governm Policy* 19:413–442. <https://doi.org/10.1068/c1v>
- Braun D, Hinze S, Hüsing B, Kuhlmann S, Menrad K, Peter V (1997) Interdisziplinäre Zentren für Klinische Forschung in der Frühphase der Bundesförderung—Vergleichende Analyse: Zwischenbericht zur Prozeßevaluation. Fraunhofer-IRB-Verl, Stuttgart
- Bräunling G, Harmsen DM (1975) Die Förderungsinstrumente der Forschungs- und Technologiepolitik, Schriftenreihe der Kommission für wirtschaftlichen und sozialen Wandel, Bd. 85, Göttingen
- Bravo-Biosca A (2019) *Experimental innovation policy*. MIT Press, Cambridge

- Brisolla S (1996) Innovation indicators for less developed countries. *Res Eval* 6:187–192. <https://doi.org/10.1093/rev/6.3.187>
- Bruno N, Kadunc M (2019) Impact pathways: tracking and communicating the impact of the European Framework Programme for research and innovation. *fteval J Res Technol Policy Eval* 62–71. <https://doi.org/10.22163/fteval.2019.330>
- Bührer S (2001) Die Nutzung von Zukunftsstudien im Rahmen der Systemevaluation der Fraunhofer-Gesellschaft—neue Ansätze der Evaluation von Forschungseinrichtungen. In: Deutsche Gesellschaft für Evaluation e.V. (ed) *Evaluation—Reformmotor oder Reformbremse?* DeGEval, Köln, pp 155–160
- Bührer S (2002) Network analysis. In: European Commission & Joanneum Reserach (ed) *EPUB RTD Evaluation Toolbox*, Sevilla, pp 163–169
- Bührer S, Kuhlmann S (1999) Evaluation of science and technology in the new Europe. In: *Proceedings of an international conference on 7 and 8 June 1999, Berlin*,
- Bührer S, Peter V (1999) *Kommunikation & Kooperation in interdisziplinären Zentren für klinische Forschung*. Fraunhofer IRB-Verlag, Stuttgart
- Bührer S, Bierhals R, Hullmann A, Studer T, Erlinghagen R (2000) *Begleitende Evaluation der Kompetenzzentren der Nanotechnologie*. Statusanalyse, Karlsruhe
- Bührer S, Bierhals R, Heinze T, Studer T, Erlinghagen R (2001) *Begleitende Evaluation der Kompetenzzentren der Nanotechnologie in der Frühphase der Bundesförderung*. Fraunhofer Institute for Systems and Innovation Research (ISI), Karlsruhe
- Bührer S, Bierhals R, Heinze T, Hullmann A, Studer T, Erlinghagen R et al (2002) *Die Kompetenzzentren der Nanotechnologie in der Frühphase der Bundesförderung : Ein Bericht der begleitenden Evaluation*. Fraunhofer ISI, Karlsruhe
- Bührer S, Edler J, Hinze S, Kuhlmann S, Braun D (2004) *Förderung interdisziplinärer Zentren für klinische Forschung: Abschlussequation und Ausblick: Prozessevaluation der Fördermaßnahme des Bundesministeriums für Bildung und Forschung “Interdisziplinäre Zentren für Klinische Forschung an den Hochschulkliniken”*. Fraunhofer IRB-Verlag, Stuttgart
- Bührer S, Wallwaey E, Seus S, Edler J (2021) Policy brief, issue 8/SIPER database, the increasing role of studies on social & environmental impacts
- Bührer S, Feidenheimer A, Frietsch R, Karaulova M, Schubert T (2022a) Report on the impact of science. Current challenges, concepts and state of the art in impact assessment
- Bührer S, Feidenheimer A, Walz R, Lindner R, Beckert B, Wallwaey E (2022b) Concepts and methods to measure societal impacts—an overview. Discussion papers innovation systems and policy analysis no. 74. Fraunhofer ISI, Karlsruhe
- Bush V (1945) *Science, the endless frontier. A report to the President by Vannevar Bush*. Director of the Office of Scientific Research and Development. United States Government Printing Office, Washington
- Cabo PG, Bijmolt THA (1992) International R&D networks: The Eureka map. *Res Eval* 2:161–168. <https://doi.org/10.1093/rev/2.3.161>
- Cañibano C, Bozeman B (2009) Curriculum vitae method in science policy and research evaluation: the state-of-the-art. *Res Eval* 18:86–94. <https://doi.org/10.3152/095820209X441754>
- Chen H-T, Rossi PH (1983) Evaluating with sense. *Eval Rev* 7:283–302. <https://doi.org/10.1177/0193841X8300700301>
- Cozzens S (2010) Social impact assessment in U.S. S&T policy. In: *Presentation to the SIAMPI workshop*, 12 Mar 2010
- Crozier M (1965) Pour une analyse sociologique de la planification française. *Revue française de sociologie* 6:147–163
- Cunningham P, Nedeva M (1999) Towards a system of continuous evaluation and monitoring for European co-operation in scientific and technical research (COST). *Res Eval* 8:142–154. <https://doi.org/10.3152/147154499781777487>
- Cunningham P, Miles I, Nedeva M, Stein J, Husson, JP, de Montgolfier P et al (1997) *COST evaluation: main report*, Brussels
- Dale AJ, Barker KE (1994) The evaluation of EUREKA: a pan-European collaborative evaluation of a pan-European collaborative technology programme. *Res Eval* 4:66–74

- Daniel H-D (1989) Ansätze zur Messung und Beurteilung des Leistungsstandes von Forschung und Technologie. *Beiträge zur Hochschulforschung* 223–231
- Da Silva CM, Henriques L (1995) R&D evaluation in Portugal. *Res Eval* 5:89–97. <https://doi.org/10.1093/rev/5.1.89>
- DeGEval—Gesellschaft für Evaluation e.V. (ed) (2016) Standards für Evaluation. Langfassung. Available online at <https://www.degeval.org/degeval-standards-alt/download/>
- Dekker W et al (1991) Report of the EUREKA assessment panel. EUREKA Secretariat
- Delahais T, Toulemonde J (2012) Applying contribution analysis: Lessons from five years of practice. *Evaluation* 18:281–293. <https://doi.org/10.1177/1356389012450810>
- Deutsch KW (1973) Zum Verständnis von Krisen und politischen Revolutionen. In: Jänicke M (ed) *Herrschaft und Krise*. VS Verlag für Sozialwissenschaften, Wiesbaden
- Diercks G, Larsen H, Steward F (2019) Transformative innovation policy: Addressing variety in an emerging policy paradigm. *Res Policy* 48:880–894. <https://doi.org/10.1016/j.respol.2018.10.028>
- Dinges M, Meyer S, Brodnik C (2020a) Key elements of evaluation frameworks for transformative R&I programmes in Europe. *fteval J Res Technol Policy Eval* 26–40. <https://doi.org/10.22163/fteval.2020.486>
- Dinges M, Wang A, Schuch K (2020b) Using the Delphi Method in evaluations—incorporating a future oriented perspective in evaluations. *fteval J Res Technol Policy Eval* 44–50. <https://doi.org/10.22163/fteval.2020.469>
- Donovan C (2011) State of the art in assessing research impact: introduction to a special issue. *Res Eval* 20:175–179. <https://doi.org/10.3152/095820211X13118583635918>
- Donovan C, Hanney S (2011) The ‘payback framework’ explained. *Res Eval* 20:181–183. <https://doi.org/10.3152/095820211X13118583635756>
- Elder J, Lo V (2004) Assessment “Future of the competence centre programmes (Kplus und Kind/net) and future of the competence centres”: executive summary. Fraunhofer Institute for Systems and Innovation Research (ISI); KMU Forschung Austria - Austrian Institute for SME Research, Karlsruhe
- Elder J, Bühler S, Lo V, Rainfurth C, Kuhlmann S, Sheikh S (2004) Assessment “Zukunft der Kompetenzzentrenprogramme (K plus und K ind/net) und Zukunft der Kompetenzzentren”. Fraunhofer Institut, Karlsruhe
- Elder J, Joly PB, Kuhlmann S, Nedeva M, Propp T, Rip A et al (2006) Understanding “fora of strategic intelligence for research and innovation”. The PRIME Forum Research Project, Report on major results, Strategic review. Manchester Institute of Innovation Research, Manchester
- Elder J, Ebersberger B, Lo V (2008) Improving policy understanding by means of secondary analyses of policy evaluation. *Res Eval* 17:175–186. <https://doi.org/10.3152/095820208X331667>
- Elder J, Amanatidou E, Berger M, Bühler S, Cunningham P, Daimer S et al (2010) INNO-appraisal. Perspectives on evaluation and monitoring. Final report: (contract number: 04377). Manchester Institute of Innovation Research, Manchester
- Elder J, Cunningham P, Gok A, Shapira P (2016) Handbook of innovation policy impact. EU-SPRI Forum on Science, Technology and Innovation Policy. Edward Elgar, Northampton, MA
- Etzioni A (1968) *The active society*, 2nd edn. Free Press, New York
- European Commission, Directorate-General for Research and Innovation, Second European report on S&T indicators 1997 – Report, Publications Office, 1997
- European Commission, Directorate-General for Research and Innovation (2006) *Creating an innovative Europe : report of the independent expert group on R&D and innovation appointed following the Hampton Court Summit*. Publications Office
- Evaluierungskommission (ed) (1998) *Systemevaluierung der Fraunhofer-Gesellschaft*. Bericht der Evaluierungskommission, München
- Fagerberg J, Martin BR, Andersen ES (2013) *Innovation studies: evolution and future challenges*. Oxford University Press, Oxford
- Fahrenkrog G, Polt W, Rojo J, Zinöcker K, Tübke A (2002) *RTD Evaluation Toolbox—assessing the socio-economic impact of RTD-Policies*, Sevilla

- Fayl G, Dumont Y, Durieux L, Karatzas I, O'Sullivan L (1998) Evaluation of research and technological development programmes: a tool for policy design. *Res Eval* 7:93–97. <https://doi.org/10.1093/rev/7.2.93>
- Fayl G, Dumont Y, O'Sullivan L, Durieux L, Karatzas I, Teuber H (1999) Overview of the EC RTD programme monitoring and evaluation system. *fteval J Res Technol Policy Eval* 8:2–5
- Feidenheimer A, Frietsch R, Schubert T, Neuhäusler P (2018) Intermediate report on the conceptual framework, project report. Big data approaches for improved monitoring of research and innovation performance and assessment of the societal impact in the Health, Demographic Change and Wellbeing Societal Challenge. Fraunhofer ISI, Karlsruhe
- Feller I (2007) Mapping the frontiers of evaluation of public-sector R&D programs. *Sci Public Policy* 34:681–690. <https://doi.org/10.3152/030234207X258996>
- Ferlie E, Ashburner L, Pettigrew A (1996) *The new public management in action*. Oxford University Press, Oxford
- Firpo T, Phipps J (2019) New running experiments in innovation and growth policy: What can we learn from recent experience? *fteval J Res Technol Policy Eval* 47:46–50. <https://doi.org/10.22163/fteval.2019.328>
- Flecha R, Radauer A, Besselaar P (2018) Monitoring the impact of EU framework programmes: expert report. Publications Office
- Freeman C, Soete L (1997) *The economics of industrial innovation*. MIT Press, Cambridge, MA
- Friedmann J (1987) *Planning in the public domain: From knowledge to action*. Princeton University Press, Princeton
- ftval-journal (2022) Ausgabe Nr. 53 *fteval Journal for Research and Technology Policy Evaluation conference proceeding: Proceedings of the revaluation conference 2021/22*
- Gabolde J (1998) New challenges for indicators in science and technology policy-making: a European view. *Res Eval* 7:99–104. <https://doi.org/10.1093/rev/7.2.99>
- Georghiou L (1995a) Research evaluation in European national science and technology systems. *Res Eval* 5:3–10. <https://doi.org/10.1093/rev/5.1.3>
- Georghiou L (1995b) Assessing the framework programmes. *Evaluation* 1:171–188. <https://doi.org/10.1177/135638909500100204>
- Georghiou L (1999) Socio-economic effects of collaborative R&D—European experiences. *J Technol Transf* 24:69–79. <https://doi.org/10.1023/A:1007724804288>
- Georghiou L, Meyer-Krahmer F (1992) Evaluation of socio-economic effects of European Community R&D programmes in the SPEAR network. *Res Eval* 2:5–15. <https://doi.org/10.1093/rev/2.1.5>
- Georghiou L, Cameron H, Stein JA, Nevada M, Janes M, Yates J, Senker J (1993) *The impact of European community policies for research and technological development upon science and technology in the United Kingdom*. HMSO, London
- Georghiou L, Rigby J, Cameron H, Vonortas NS, Prastacos G, Spanos Y, Guy K (2002) *Assessing the socio-economic impacts of the framework programme*. PREST, Manchester
- Geuna A (2001) The changing rationale for European university research funding: are there negative unintended consequences? *J Econ Issues* 35:607–632
- Grupp H, Hohmeyer O (1986) A technometric model for the assessment of technological standards and their application to selected technology-intensive products. *Technol Forecast Soc Change* 30:123–137
- Grupp H, Schmoch U, Kuntze U (1995) New technology indicators for the evaluation of research and development programmes. In: Becher G, Kuhlmann S (eds) *Evaluation of technology policy programmes in Germany*. Springer, Dordrecht, pp 243–284
- Guba EG, Lincoln YS (1989) *Fourth generation evaluation*, 13th edn. Sage, Newbury Park
- Guy K, Polt W (1999) Strategic options for the evaluation of the R&D programmes of the European Union. *fteval J Res Technol Policy Eval* 8:5–9
- Hage J, Jordan G, Mote J (2007) A theory-based innovation systems framework for evaluating diverse portfolios of research, part two: macro indicators and policy interventions. *Sci Public Policy* 34:731–741. <https://doi.org/10.3152/030234207X265385>

- Hall PA, Soskice DW (2001) Varieties of capitalism: the institutional foundations of comparative advantage. Oxford University Press, Oxford
- Hicks D (1991) A cautionary view of co-citation analysis. *Res Eval* 1:31–36. <https://doi.org/10.1093/rev/1.1.31>
- Hicks D (2012) Performance-based university research funding systems. *Res Policy* 251–261
- Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I (2015) Bibliometrics: The Leiden Manifesto for research metrics. *Nature* 520:429–431. <https://doi.org/10.1038/520429a>
- Hills PV, Dale AJ (1995) Research and technology evaluation in the United Kingdom. *Res Eval* 5:35–44. <https://doi.org/10.1093/rev/5.1.35>
- Holland D, Kuhlmann S (1995) Systemwandel und industrielle Innovation: Studien zum technologischen und industriellen Umbruch in den neuen Bundesländern. *Technik, Wirtschaft und Politik*, vol 16. Physica-Verlag HD, Heidelberg
- Internationale Kommission (1999) Forschungsförderung in Deutschland. Bericht der internationalen Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft. Volkswagen-Stiftung, Hannover
- JIIP (2018) Mission-oriented research and innovation : assessing the impact of a mission-oriented research and innovation approach: Final report. Publications Office, Brussels
- Joly P-B, Matt M (2017) Towards a new generation of research impact assessment approaches. *J Technol Transf.* <https://doi.org/10.1007/s10961-017-9601-0>
- Joly P-B, Gaunand A, Colinet L, Larédo P, Lemarié S, Matt M (2015) ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. *Res Eval* 24:440–453. <https://doi.org/10.1093/reseval/rvv015>
- Jongbloed B, Lepori B (2015) The funding of research in higher education: mixed models and mixed results. In: Huisman J, De Boer H, Dill DD, Souto-Otero M (eds) *The Palgrave international handbook of higher education policy and governance*. Palgrave Macmillan, London, pp 439–462
- Jongbloed B, Vossensteyn H (2001) Keeping up performances: an international survey of performance-based funding in higher education. *J High Educ Policy Manag* 23:127–145. <https://doi.org/10.1080/13600800120088625>
- Kastrinos N, Katsoulacos Y (1995) Towards a national system of research evaluation in Greece. *Res Eval* 5:63–68. <https://doi.org/10.1093/rev/5.1.63>
- Kingsbury N (1986) Coming together: evaluation network and evaluation research society share common business agendas at Evaluation '85 leading to the American Evaluation Association. *Eval Pract* 7:107–110
- Kohlweg K (2019) Standards der Evaluierung in der Forschungs-, Technologie-, und Innovationspolitik. Technischer Bericht. fteval—Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung, Vienna. Available online at <https://repository.fteval.at/387/>
- Koschatzky K (1996) Technology transfer, innovation networking and regional development: Paper presented at the RESTPOR '96 conference on “global comparison of regional RTD & innovation strategies for development and cohesion”, 19–21 Sept 1996, Brussels. Arbeitspapier Regionalforschung, Nr. 4. ISI, Karlsruhe
- Koschatzky K (2000) The regionalisation of innovation policy in Germany—theoretical foundations and recent experience. Working Papers Firms and Regions, R1/2000, Karlsruhe
- Krauch H (1970) *Prioritäten für die Forschungspolitik*. Hanser, München
- Kroll H, Hansmeier H, Hufnagl M (2022) Productive interactions in basic research an enquiry into impact pathways at the DESY synchrotron. *Technological forecasting and social change*, vol 175
- Krull W (1992) Perspektiven für Wissenschaft und Forschung in den neuen Lndern. *Naturwissenschaften* 79:49–59. <https://doi.org/10.1007/BF01131803>
- Kuhlmann S (1995) Patterns of science and technology policy evaluation in Germany. *Res Eval* 5:23–33. <https://doi.org/10.1093/rev/5.1.23>
- Kuhlmann S (1998a) Moderation of policy-making? *Evaluation* 4:130–148. <https://doi.org/10.1177/13563899822208491>

- Kuhlmann S (1998b) *Politikmoderation: Evaluationsverfahren in der Forschungs- und Technologiepolitik*. Nomos Verlagsgesellschaft, Baden-Baden
- Kuhlmann S (2009) Evaluation von Forschungs- und Innovationspolitik in Deutschland: Stand und Perspektiven. In: Widmer T, Beywl W, Fabian C (eds) *Evaluation: Ein systematisches Handbuch*. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 283–294
- Kuhlmann S, Arnold E (2001) RCN in the Norwegian research and innovation system: background report no 12 in the evaluation of the Research Council of Norway
- Kuhlmann S, Holland D (1995) Erfolgsfaktoren der wirtschaftsnahen Forschung. In: *Technik, Wirtschaft und Politik, Schriftenreihe des Fraunhofer-Instituts für Systemtechnik und Innovationsforschung (ISI)*, vol 15. Physica-Verlag HD; Imprint; Physica, Heidelberg
- Kuhlmann S, Holland D, Grupp H (1995) Evaluation von Technologiepolitik in Deutschland: Konzepte, Anwendung, Perspektiven. *Technik, Wirtschaft und Politik*, vol 12. Physica-Verlag, Heidelberg
- Kuhlmann S, Boekholt P, Georghiou L, Guy K, Héraud J-A, Larédo P, Lemola T, Loveridge D, Luukkonen T, Polt W, Rip A, Sanz-Menendez L, Smits R (1999) Improving distributed intelligence in complex innovation systems: final report of the Advanced Science & Technology Policy Planning Network (ASTPP), Karlsruhe
- Kulicke M (1993) Chancen und Risiken Junger Technologieunternehmen: Ergebnisse des Modellversuchs “Förderung technologieorientierter Unternehmensgründungen” des Bundesforschungsministeriums. *Technik, Wirtschaft und Politik*, vol 4. Physica-Verlag, Heidelberg
- Kulicke M (2002) Multi-Akteurs-/Multi-Maßnahmen-Wettbewerbe und ihre Anforderungen an die wissenschaftliche Evaluation—Das Beispiel EXIST. In: *Wissenschaft und Wirtschaft im regionalen Gründungskontext*, pp 45–66
- Kulicke M, Görisch J, Tönnissen F (2000) EXIST Existenzgründer aus Hochschulen. Netzwerke für innovative Unternehmens-gründungen, Bonn
- Kulicke M, Görisch J, Stahlecker T (2002) Erfahrungen aus EXIST. Querschau über die einzelnen Projekte, Bonn
- Landon T, Hochreiter H (2022) Randomised controlled trials and other experimental approaches in the Austrian Research Promotion Agency. *fteval J Res Technol Policy Eval* 160–168. <https://doi.org/10.22163/fteval.2022.554>
- Larédo P, Callon M (1990) L’impact des programmes communautaires sur le tissu scientifique et technique français: rapport du Centre de sociologie de l’innovation de l’Ecole des mines de Paris. Documentation française
- Larédo P, Mustar P (1995) France, the guarantor model and the institutionalisation of evaluation. *Res Eval* 5:11–21. <https://doi.org/10.1093/rev/5.1.11>
- Leko N et al (2016) How policy formation changes with evidence: Trends toward evidence-based policy in the UK, as verified by randomised controlled trials, and some implications for Japan (Mitsubishi UFJ Research and Consulting Policy Research Report). Available online at UFJ Research and Consulting Policy Research Report. [http://www.murc.jp/uploads/2016/02/seiken\\_160212\\_1.pdf](http://www.murc.jp/uploads/2016/02/seiken_160212_1.pdf). Checked on 1 Mar 2017
- Lepori B, Probst C (2009) Using curricula vitae for mapping scientific fields: a small-scale experience for Swiss communication sciences. *Res Eval* 18:125–134. <https://doi.org/10.3152/095820209X441772>
- Lindner R, Edler J, Daimer S (2024) Understanding paradigm change in science, technology and innovation policy: Between science push and policy pull. In: Edler J, Walz R (eds) *Systems and innovation research in transition: Research questions and trends in historical perspective*. Springer, Heidelberg
- Lipset SM (1959) Some social requisites of democracy: economic development and political legitimacy. *Am Polit Sci Rev* 53:69–105. <https://doi.org/10.2307/1951731>
- Luukkonen T (1991) Citation indicators and peer review: their time-scales, criteria of evaluation, and biases. *Res Eval* 1:21–30. <https://doi.org/10.1093/rev/1.1.21>
- Luukkonen T, Niskanen P (1998) Learning through collaboration: Finnish participation in EU framework programmes. VTT Technical Research Centre of Finland

- Magro E, Wilson JR (2013) Complex innovation policy systems: Towards an evaluation mix. *Res Policy* 42:1647–1656. <https://doi.org/10.1016/j.respol.2013.06.005>
- Mandel E (1986) In defence of socialist planning. *New Left Rev* 159:5–37
- Martin BR (1996) Technology foresight: Capturing the benefits from science-related technologies. *Res Eval* 6:158–168. <https://doi.org/10.1093/rev/6.2.158>
- Martinsen R (1992) Theorien politischer Steuerung—auf der Suche nach dem dritten Weg. In: Grimmer K, Häusler J, Kuhlman S, Simonis G (eds) *Politische Techniksteuerung*. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 51–73
- Mayne J (2001) Addressing attribution through contribution analysis: using performance measures sensibly. *Can J Program Eval* 16:1–24
- Mayne J (2012) Contribution analysis: Coming of age? *Evaluation* 18:270–280. <https://doi.org/10.1177/1356389012451663>
- Mayntz R (1979) Public bureaucracies and policy implementation. *Int Soc Sci J* 31:633–645
- Mayntz R (1980) *Implementation politischer Programme: Empirische Forschungsberichte*. Verlagsgruppe Athenäum, Hain, Scriptor, Hanstein, Königsstein/Ts
- Mayntz R, Scharpf FW (1975) *Policy making in the German Federal Bureaucracy*. Elsevier, Amsterdam
- Mazzucato M (2018) *Mission-oriented research & innovation in the European Union: A problem-solving approach to fuel innovation-led growth*. Publications Office of the European Union, Brussels
- Meyer N, Bühner S (2014) Impact evaluation of the Erwin Schrödinger fellowships with return phase. Final report for the Austrian Science Fund (FWF), Vienna
- Meyer-Krahmer F (1981) The present status and problems of impact research in technology policy: A case study on the federal program for funding research and development personnel in Germany. *Res Policy* 10:356–366
- Meyer-Krahmer F, Reiss T (1992) Ex ante evaluation and technology assessment—Two emerging elements of technology policy evaluation. *Res Eval* 2:47–54. <https://doi.org/10.1093/rev/2.1.47>
- Meyer-Krahmer F, Gielow G, Kuntze U (1983) Impacts of government incentives towards industrial innovation: An analysis of the federal programme funding R&D personnel in the Federal Republic of Germany. *Res Policy* 12:153–169
- Mickwitz P, Neij L, Johansson M, Benner M, Sandin S (2021) A theory-based approach to evaluations intended to inform transitions toward sustainability. *Evaluation* 27:281–306. <https://doi.org/10.1177/1356389021997855>
- Miles I, Cunningham P, Cox D, Crowe C, Malik K (2006) *Smart innovation: supporting the monitoring and evaluation of innovation programmes*. European Commission: Enterprise and Industry
- Moed H, Burger WJM, Frankfort JG, van Raan AFJ (1985) The use of bibliometric data for the measurement of university research performance. *Res Policy* 14:131–149
- Molas-Gallart J, Davies A (2006) Toward theory-led evaluation. *Am J Eval* 27:64–82. <https://doi.org/10.1177/1098214005281701>
- Molas-Gallart J, Tang P (2011) Tracing ‘productive interactions’ to identify social impacts: an example from the social sciences. *Res Eval* 20:219–226. <https://doi.org/10.3152/095820211X12941371876706>
- Molas-Gallart J, Boni A, Giachi S, Schot J (2021) A formative approach to the evaluation of transformative innovation policies. *Res Eval*. <https://doi.org/10.1093/reseval/rvab016>
- Morton S (2015) Progressing research impact assessment: A ‘contributions’ approach. *Res Eval* 24:405–419. <https://doi.org/10.1093/reseval/rvv016>
- Muhonen R, Benneworth P, Olmos-Peñuela J (2019) From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact. *Res Eval*. <https://doi.org/10.1093/reseval/rvz003>
- Naschold F (1969) *Systemsteuerung*. Geschichte und Gegenwart, vol 2. Kohlhammer, Stuttgart
- Noyons ECM, van Raan AFJ (1996) Actor analysis in neural network research: The position of Germany. *Res Eval* 6:133–142. <https://doi.org/10.1093/rev/6.2.133>

- OECD (ed) (1997) Policy evaluation in innovation and technology: towards best practices. OECD, Paris
- OECD (2002) Benchmarking industry-science relationships. Organisation for Economic Co-operation and Development, Paris
- OECD Proceedings (1999) Boosting innovation: the cluster approach, Paris
- Ormala E (1993) Evaluation of EUREKA industrial and economic effects, Brussels
- Padilla P (2016) Policy learning through strategic intelligence: The American Small Business Innovation Research Program (SBIR) and British Small Business Research Initiative (SBRI). University of Twente, Enschede
- Patton MQ (1997) Utilization-focused evaluation: the new century text. Sage, Thousand Oak
- Patton MQ (2011) Developmental evaluation. Guilford Press, New York
- Patton MQ (2019) Blue marble evaluation: premises and principles. Guilford, New York
- Pavitt K (1998) Do patents reflect the useful research output of universities? *Res Eval* 7:105–111. <https://doi.org/10.1093/rev/7.2.105>
- Picht G (1964) Die deutsche Bildungskatastrophe: Analyse und Dokumentation. Walter, Freiburg im Breisgau
- Pinar M, Horne TJ (2022) Assessing research excellence: evaluating the research excellence framework. *Res Eval* 31:173–187. <https://doi.org/10.1093/reseval/rvab042>
- Polt W (2003) New developments in evaluation methods and strategy at the European level—a short review of recent projects (ASIF, EPUB). *fteval J Res Technol Policy Eval* 11–12
- Polt W, Ploder M, Breitfuss M, Daimer S, Jackwerth T, Zielinski A (2021) Politikstile und Politikinstrumente in der F&I-Politik. Studien zum deutschen Innovationssystem. EFI, Berlin
- Pressman JL, Wildavsky A (1984) Implementation: How great expectations in Washington are dashed in Oakland; or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation, vol 708. University of California Press, Berkeley
- Reale E, Nedeva M, Duncan T, Primeri E (2014) Evaluation through impact: a different viewpoint. *fteval J Res Technol Policy Eval* 36–41
- Reale E, Avramov D, Canhial K, Donovan C, Flecha R, Holm P, Larkin C, Lepori B, Mosoni-Fried J, Oliver E, Primeri E, Puigvert L, Scharnhorst A, Schubert A, Soler M, Soòs S, Sordé T, Travis C, van Horik R (2018) A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. *Res Eval* 27:298–308. <https://doi.org/10.1093/reseval/rvx025>
- Reger G, Kuhlman S (1995) European technology policy in Germany. Technology, innovation, and policy, vol 2. Physica; Springer series, Heidelberg
- Reger G, Balthasar A, Bühler S, Bättig C (1998) Switzerland's participation in the European RTD framework programmes: a win-win game? *Technovation* 18:425–438. [https://doi.org/10.1016/s0166-4972\(98\)00032-7](https://doi.org/10.1016/s0166-4972(98)00032-7)
- Removille J, Clarysse B (1999) Intra-European scientific co-operation: measuring policy impact. *Res Eval* 8:99–109. <https://doi.org/10.3152/147154499781777586>
- Rip A, van der Meulen BJR (1995) The patchwork of the Dutch evaluation system. *Res Eval* 5:45–53. <https://doi.org/10.1093/rev/5.1.45>
- Roelandt T, van der Wiel H (2020) The long-term impact of Dutch innovation vouchers: Back to the future with randomised controlled trials. <https://www.innovationgrowthlab.org/blog/long-term-impact-dutch-innovation-vouchers-back-future-randomised-controlled-trials>
- Rogers JD, Jordan G (2010) New research evaluation frameworks and methods for systems level learning: introduction to a special section. *Res Eval* 19:235–237. <https://doi.org/10.3152/095820210X12827366906526>
- Rojo J (2003) Evaluation of RTD policy foundations: the socio-economic dimension. *fteval J Res Technol Policy Eval* 18–24
- Ruegg R (2006) Bridging from project case study to portfolio analysis in a public R&D program a framework for evaluation and introduction to a composite performance rating system (NIST GCR 06-891)

- Ruegg RT (2007) Quantitative portfolio evaluation of US federal research and development programs. *Sci Public Policy* 34:723–730. <https://doi.org/10.3152/030234207X259021>
- Ruegg R, Feller I (2003) A toolkit for evaluating public R&D investment models, methods, and findings from ATP's first decade (NIST GCR 03–857)
- Sandström U (2009) Combining curriculum vitae and bibliometric analysis: mobility, gender and research performance. *Res Eval* 18:135–142. <https://doi.org/10.3152/095820209X441790>
- Sanz-Menéndez L (1995) Research actors and the state: research evaluation and evaluation of science and technology policies in Spain. *Res Eval* 5:79–88. <https://doi.org/10.1093/rev/5.1.79>
- Schmoch U (1999) Impact of international patent applications on patent indicators. *Res Eval* 8:119–131. <https://doi.org/10.3152/147154499781777540>
- Schmoch U, Grupp H, Reiss T, Strauss E (1991) Monitoring science-based biotechnology by means of patent indicators. *Res Eval* 1:61–68. <https://doi.org/10.1093/rev/1.2.61>
- Schofer E, Meyer JW (2005) The worldwide expansion of higher education in the twentieth century. *Am Sociol Rev* 70:898–920. <https://doi.org/10.1177/000312240507000602>
- Schön DA, Rein M (1994) *Frame reflection: Toward the resolution of intractable policy controversies*. Basic Books, New York
- Schot J, Steinmueller WE (2018) Three frames for innovation policy: R&D, systems of innovation and transformative change. *Res Policy* 47:1554–1567. <https://doi.org/10.1016/j.respol.2018.08.011>
- Scriven M (1991) *Evaluation Thesaurus*, 4th edn. Sage, Newbury Park
- Seus S, Bühler S (2021) How to evaluate a transition-oriented funding programme? Lessons learned from the evaluation of FONA, the German framework programme to promote sustainability research. *fteval J Res Technol Policy Eval* 10–18. <https://doi.org/10.22163/fteval.2021.514>
- Seus S, Heckl E, Bühler S (2016) Evaluation of the START programme and the Wittgenstein Award. Available online at [https://zenodo.org/record/50610/files/Eval-START-Witt\\_final\\_report.pdf](https://zenodo.org/record/50610/files/Eval-START-Witt_final_report.pdf). Checked on 13 Feb 2017
- Shapira P, Kuhlmann S (eds) (2003) *Learning from science and technology policy evaluation: Experiences from the United States and Europe*. Edward Elgar, Cheltenham
- Silvani A, Sirilli G (1995) R&D evaluation in Italy: A science and technology policy view. *Res Eval* 5:69–77. <https://doi.org/10.1093/rev/5.1.69>
- Spaapen J, van Drooge L (2011) Introducing ‘productive interactions’ in social impact assessment. *Res Eval* 20:211–218. <https://doi.org/10.3152/095820211X12941371876742>
- Spaapen J, van Drooge L, Propp T, van der Meulen B, Shinn T, Marcovich A (2011) *Social impact assessment methods for research and funding instruments through the study of productive interactions (SIAMPI)*. Final report on social impacts of research. Royal Netherlands Academy of Arts and Sciences, Amsterdam
- Steiner M, Sturn D (1995) Elements of evaluation of science and technology policy in Austria. *Res Eval* 5:98–108. <https://doi.org/10.1093/rev/5.1.98>
- Stern E (1993) Ongoing and participative evaluation: Purpose, design and role in the evaluation of a large-scale R&D programme. *Res Eval* 3:75–82. <https://doi.org/10.1093/rev/3.2.75>
- Stockmann R (ed) (2000) *Evaluationsforschung*. Leske+Budrich, Leverkusen
- Strassheim H, Kettunen P (2014) When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evid Policy* 10:259–277. <https://doi.org/10.1332/174426514X13990433991320>
- Tomizawa H, Niwa F (1996) Evaluating overall national science and technology activity: general indicator of science and technology (GIST) and its implications for S&T policy. *Res Eval* 6:83–92. <https://doi.org/10.1093/rev/6.2.83>
- van Raan AF (ed) (1988) *Handbook of quantitative studies of science and technology*. Elsevier, Amsterdam
- van Raan AFJ (1993) Advanced bibliometric methods to assess research performance and scientific development: basic principles and recent practical applications. *Res Eval* 3:151–166. <https://doi.org/10.1093/rev/3.3.151>
- van Steen J (1995) S&T indicators in science policy: how can they matter? *Res Eval* 5:161–166. <https://doi.org/10.1093/rev/5.2.161>

- van Steen J, Eijffinger M (1998) Evaluation practices of scientific research in the Netherlands. *Res Eval* 7:113–122. <https://doi.org/10.1093/rev/7.2.113>
- Vonortas NS, Stampfer M, Zinöcker K (2007) Evidence never lies: introduction to a special issue on new frontiers in evaluation. *Sci Public Policy* 34:679–680. <https://doi.org/10.3152/030234207X273548>
- Wagner RM, Jordan GB (2009) Moving towards impact when evaluating research programs: introduction to a special section. *Res Eval* 18:339–342. <https://doi.org/10.3152/095820209X483073>
- Warta K, Philipp S (2014) Was bringt die Plattform fteval und zu was hat sie es gebracht? Rückblick und Ergebnisse einer Mitgliederbefragung. *fteval J Res Technol Policy Eval* 15–24
- Weber KM, Polt W (2014) Assessing mission-orientated R&D programs: combining foresight and evaluation. *fteval J Res Technol Policy Eval* 11–15
- Weingart P, Sehringer R, Winterhager M (1988) Bibliometric indicators for assessing strengths and weaknesses of West German science. In: van Raan AF (ed) *Handbook of quantitative studies of science and technology*. Elsevier, Amsterdam, pp 391–430
- Weiss CH (1993) Where politics and evaluation research meet. *Eval Pract* 14:93–106
- Weiss CH (1997) Theory-based evaluation: past, present, and future. *New Dir Eval* 41–55
- Widmer T (1996) *Meta-Evaluation: Kriterien zur Bewertung von Evaluationen*. Haupt, Bern
- Wilsdon J, Allen L, Belfiore E, Campbell P, Curry S, Hill S et al (2015) *The metric tide: report of the independent review of the role of metrics in research assessment and management*. Checked on 22 May 2022
- Wilsdon J, Bar-Ilan J, Frodeman R, Lex E, Peters I, Wouters P (2017) *Next-generation metrics: responsible metrics and evaluation for open science*. Report. European Commission, Brussels
- Wissenschaftsrat (ed) (2000) *Systemevaluation der Blauen Liste—Stellungnahme des Wissenschaftsrates zum Abschluss der Bewertung der Einrichtungen der Blauen Liste*, Köln
- Wissenschaftsrat (ed) (2001) *Systemevaluation der HGF—Stellungnahme des Wissenschaftsrates zur Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren*, Köln
- Wittmann F, Hufnagl M, Roth F, Lindner R, Kroll H (2022) *A framework for formative evaluation and impact assessment of mission-oriented innovation policies*. Final report of the Scientific Support Action to the German High-Tech Strategy 2025—volume 2. Checked on 17 May 2022
- Wollmann H (1998) *Evaluation research and politics: Between a science-driven and a pluralist controversy-responsive policy-making model*. Potential and limitations. Paper given to the conference of the European Evaluation Society (EES). Rom
- Worthen BR, Sanders JR, Fitzpatrick JL (1997) *Program evaluation: alternative approaches and practical guidelines*, 2nd edn. Longman, New York
- Zellner C (2002) Evaluating the social economic benefits of publicly funded basic research via scientists' career mobility. *Res Eval* 11:27–35. <https://doi.org/10.3152/147154402781776970>
- Zulueta MA, Bordons M (1999) A global approach to the study of teams in multidisciplinary research areas through bibliometric indicators. *Res Eval* 8:111–118

**Susanne Bühner** is coordinator of the Business Unit “Societal Change and Innovation” at the Fraunhofer Institute for Systems and Innovation Research. Her research areas are the development and implementation of sophisticated evaluation and impact analyses, the conceptualisation of impact-oriented monitoring systems, the topic of women in research and innovation systems (with special consideration of Gendered Innovations) as well as questions of responsible research and innovation.

**Jakob Edler** is Managing Director of the Fraunhofer Institute for Systems and Innovation Research ISI since October 2018. The Anglo-German is also Professor of Innovation Policy and Strategy at the Manchester Institute of Innovation Research (MIOIR), which he led from 2011 to September 2018. His research focuses on analysing and conceptualising governance, policy and management of innovation. J. Edler is Speaker of the Fraunhofer Group Innovation Research, member of the German Science Council, the German Academy of Science and Engineering (Acatech) and the Royal Society for the Encouragement of Arts, Manufactures and Commerce (UK).

**Stefan Kuhlmann** is a political scientist and emeritus professor “Science, Technology and Society” (STS) at the University of Twente (Netherlands). He works on research and technological innovation as social and political processes, focusing on governance and politics, and he publishes widely in the field of research and innovation policy studies. Prior he was professor of Innovation Policy at Utrecht University (Netherlands, 2001–2006), held leading positions at the Fraunhofer Institute for Systems and Innovation Research (Germany 1988–2006) and was researcher at the University of Kassel (Germany, 1979–1988). He graduated in History and Political Science from the University of Marburg (Germany, 1978).

**Sarah Seus** is a social scientist and a researcher in the Competence Centre Policy and Society of the Fraunhofer Institute for Systems and Innovation Research ISI. Her research areas are the evaluation of science and innovation policies and specifically research on methodologies for impact evaluation (incl. societal impacts) of research funding; the analysis of STI policies and governance at national and EU level; the analysis of transition processes within the science system (e.g. trans-disciplinary research) and questions of Responsible Research and Innovation (RRI).

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

